

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 11:46:04 ; Search time 1300.69 Seconds
(without alignments)
253.213 Million cell updates/sec

Title: US-09-101-423B-8
Perfect score: 47
Sequence: 1 agctataatgcggccgata.....tgatgcgcgcgaagcttgga 47

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues 15983484
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries.

Database :

EST: *
1: gb_estl1: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_est9: *
10: gb_est10: *
11: gb_est11: *
12: gb_est12: *
13: gb_est13: *
14: gb_est14: *
15: gb_est15: *
16: gb_est16: *
17: gb_est17: *
18: gb_est18: *
19: gb_est19: *
20: gb_est20: *
21: gb_est21: *
22: gb_est22: *
23: gb_est23: *
24: gb_est24: *
25: gb_est25: *
26: gb_est26: *
27: gb_est27: *
28: gb_est28: *
29: gb_est29: *
30: gb_est30: *
31: gb_est31: *
32: gb_est32: *
33: gb_est33: *
34: gb_est34: *
35: gb_est35: *
36: gb_est36: *
37: gb_est37: *
38: gb_est38: *
39: gb_est39: *
40: gb_est40: *
41: em_estba: *
42: em_estfun: *
43: em_esthum1: *

44: em_esthum2: *
45: em_esthum3: *
46: em_esthum4: *
47: em_esthum5: *
48: em_esthum6: *
49: em_esthum7: *
50: em_esthum8: *
51: em_esthum9: *
52: em_esthum10: *
53: em_esthum11: *
54: em_esthum12: *
55: em_esthum13: *
56: em_esthum14: *
57: em_esthum15: *
58: em_esthum16: *
59: em_esthum17: *
60: em_esthum18: *
61: em_esthum19: *
62: em_esthum20: *
63: em_estin1: *
64: em_estin2: *
65: em_estin3: *
66: em_estin4: *
67: em_estov1: *
68: em_estov2: *
69: em_estpl1: *
70: em_estpl2: *
71: em_estpl3: *
72: em_estpl4: *
73: em_estpl5: *
74: em_estrol: *
75: em_estro2: *
76: em_estro3: *
77: em_estro4: *
78: em_estro5: *
79: em_estro6: *
80: em_estro7: *
81: em_estro8: *
82: em_estro9: *
83: em_estro10: *
84: em_estro11: *
85: em_estro12: *
86: em_estro13: *
87: gb_est41: *
88: gb_est42: *
89: gb_est43: *
90: gb_est44: *
91: gb_est45: *
92: gb_est46: *
93: gb_est47: *
94: gb_est48: *
95: gb_est49: *
96: gb_est50: *
97: gb_est51: *
98: gb_est52: *
99: gb_est53: *
100: gb_est54: *
101: gb_est55: *
102: gb_est56: *
103: gb_est57: *
104: gb_est67: *
105: gb_est68: *
106: gb_est69: *
107: gb_est70: *
108: gb_est71: *
109: gb_est72: *
110: gb_est73: *
111: gb_est74: *
112: em_esthum21: *
113: em_esthum22: *
114: em_esthum23: *
115: em_estom1: *
116: em_estom2: *

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp19:*
121: em_estp20:*
122: em_estp21:*
123: em_estp22:*
124: em_estp23:*
125: em_estp24:*
126: em_estp25:*
127: em_estp26:*
128: em_estp27:*
129: em_estp28:*
130: em_estp29:*
131: em_estp30:*
132: em_estp31:*
133: em_estp32:*
134: em_estp33:*
135: em_estp34:*
136: em_estp35:*
137: em_estp36:*
138: em_estp37:*
139: em_estp38:*
140: em_estp39:*
141: em_estp40:*
142: em_estp41:*
143: em_estp42:*
144: em_estp43:*
145: em_estp44:*
146: em_estp45:*
147: em_estp46:*
148: em_estp47:*
149: em_estp48:*
150: em_estp49:*
151: em_estp50:*
152: em_estp51:*
153: em_estp52:*
154: em_estp53:*
155: em_estp54:*
156: em_estp55:*
157: em_estp56:*
158: em_estp57:*
159: em_estp58:*
160: em_estp59:*
161: em_estp60:*
162: em_estp61:*
163: em_estp62:*
164: em_estp63:*
165: em_estp64:*
166: em_estp65:*
167: em_estp66:*
168: em_estp67:*
169: em_estp68:*
170: em_estp69:*
171: em_estp70:*
172: em_estp71:*
173: em_estp72:*
174: em_estp73:*
175: em_estp74:*
176: em_estp75:*
177: em_estp76:*
178: em_estp77:*
179: em_estp78:*
180: em_estp79:*
181: em_estp80:*
182: em_estp81:*
183: em_estp82:*
184: em_estp83:*
185: em_estp84:*
186: em_estp85:*
187: em_estp86:*
188: em_estp87:*
189: em_estp88:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	23.4	49.8	313	18	AI283439	AI283439 qh93e10.x
2	23.4	49.8	520	4	AA236621	AA236621 zs43g09.r
C 3	23.2	49.4	558	163	AQ841408	AQ841408 T136718b
C 4	23.2	49.4	1054	134	BE036773	BE036773 MP05D02 M
C 5	22.4	47.7	705	97	AW977483	AW977483 EST389592
C 6	22.2	47.2	528	134	BE062593	BE062593 QV1-BT026
C 7	22.2	47.2	625	21	AI514427	AI514427 LD40932.5
8	22	46.8	267	14	AA952587	AA952587 TENS1771
9	22	46.8	313	23	AI667900	AI667900 TENG0871
C 10	22	46.8	447	17	AA063113	AA063113 zf68e10.s
C 11	22	46.8	475	17	AI187524	AI187524 EST316 Ma
C 12	22	46.8	892	109	BE535565	BE535565 601059673
C 13	21.8	46.4	436	37	AV624476	AV624476 AV624476
C 14	21.8	46.4	558	170	AZ342872	AZ342872 LM0076E01
C 15	21.8	46.4	610	87	AW181098	AW181098 MGA0062r
C 16	21.8	46.4	640	87	AW181091	AW181091 MGA0032r
C 17	21.6	46.0	541	159	AQ578848	AQ578848 nbx00093F
C 18	21.6	46.0	614	169	AZ240289	AZ240289 RPCT-23-3
19	21.6	46.0	650	7	AA438826	AA438826 LUL3184.5
20	21.6	46.0	677	138	BE976052	BE976052 bs46f11.x
21	21.6	46.0	707	97	AW940801	AW940801 GH20963.3
C 22	21.4	45.5	288	40	AW144909	AW144909 707013F10
C 23	21.4	45.5	412	89	AW331456	AW331456 707013F10
C 24	21.4	45.5	482	40	AW147159	AW147159 707013F10
C 25	21.4	45.5	947	193	CNS05600	AL336561 Tetraodon
26	21.2	45.1	263	36	AV425024	AV425024 AV425024
27	21.2	45.1	266	36	AV426457	AV426457 AV426457
28	21.2	45.1	331	89	AW330028	AW330028 TENU4569
29	21.2	45.1	403	36	AV409533	AV409533 AV409533
30	21.2	45.1	463	36	AV422473	AV422473 AV422473
31	21.2	45.1	554	3	AA196000	AA196000 zp99f07.s
32	21.2	45.1	571	160	AQ615803	AQ615803 HS_5145_B
33	21.2	45.1	813	110	BE640770	BE640770 Cr12_1_A1
34	21.2	45.1	912	97	AW925378	AW925378 HVSMBg000
35	21	44.7	220	91	AW491261	AW491261 UI-M-BH3-
36	21	44.7	244	137	BE943769	BE943769 UI-M-BH3-
37	21	44.7	272	91	AW491371	AW491371 UI-M-BH3-
C 38	21	44.7	351	107	BE363902	BE363902 P11_10_G0
C 39	21	44.7	352	24	AI768069	AI768069 wi46d08.x
C 40	21	44.7	366	143	N77089	N77089 YV51608.r1
C 41	21	44.7	372	14	AB009130	AB009130 AB009130
C 42	21	44.7	391	36	AV420471	AV420471 AV420471
C 43	21	44.7	433	40	AW144910	AW144910 707013F10
C 44	21	44.7	448	110	BE647694	BE647694 UI-M-BH1-
C 45	21	44.7	463	39	AW077673	AW077673 fj36b05.y.

ALIGNMENTS

RESULT 1
AI283439
LOCUS
DEFINITION
qh93e10.x1 Soares_NFL_T_GBC.s1 Homo sapiens cDNA clone
IMAGE:1854570 3' similar to TR:P76904 P76904 SIMILAR TO ;, mRNA
sequence.
AI283439
ACCESSION
VERSION
KEYWORDS
AI283439.1 GI:3921672
EST
23-NOV-1998

Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=QV1-BT0260-281
 099-023-c01&t3=1999-10-28&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 9
 High quality sequence stop: 426.

FEATURES

source

Location/Qualifiers
 1..528
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT0260"
 /dev_stage="Adult"

/note="Organ: breast; Vector: puc18; Site: 1: SmaI; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

101 a 180 c 152 g 94 t 1 others

BASE COUNT
 ORIGIN

Query Match 47.28; Score 22.2; DB 134; Length 528;
 Best Local Similarity 77.18; Pred. No. 96;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 7 aatgcgcccgcatttcgctgctgacgcccgaag 41

Db 332 AATCGCGCCGACCTTCGCGAGAACGCGCAGCAAG 298

RESULT 7

AI514427/c

LOCUS

DEFINITION LD40932. Sprime LD Drosophila melanogaster embryo pOT2 Drosophila melanogaster cDNA clone LD40932 Sprime similar to U13637: y1

FBgn0004649 PID:g535346 SWISS-PROT:P98163, mRNA sequence.

AI514427

AI514427.1 GI:4418489

EST

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 625)

Harvey,D., Hong,I., Evans-Holm,M., Pendleton,J., Su,C., Brokstein

,P., Lewis,S. and Rubin,G.M.

BDGP/HMI Drosophila EST Project

Unpublished (1997)

Contact: Harvey, D.

G. M. Rubin-Molecular and Cell Biology

University of California Berkeley

539 LSA, Berkeley, CA 94720-3200, USA

Fax: 510 643 9947

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

Plate: 409 row: C column: 8

High quality sequence stop: 507.

Location/Qualifiers

1..625

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="LD40932"

/clone_lib="LD Drosophila melanogaster embryo pOT2"

/sex="male and female"

/dev_stage="0 to 24 hours mixed stage embryonic"

/lab_host="XLI Blue"

/note="Organ: embryo; Vector: pOT2; Site: 1: EcoRI; Site: 2: XhoI; Sized fractionated cDNAs were directly ligated into

BASE COUNT
 ORIGIN

Query Match 47.28; Score 22.2; DB 21; Length 625;
 Best Local Similarity 69.88; Pred. No. 97;
 Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 5 ataatgcgcccgcatttcgctgctgacgcccgaagcttggga 47

Db 349 ACAATGCGCGGCACATCGCTCCCGCAAAACTGCACGCTTGA 307

RESULT 8

AA952587

LOCUS

DEFINITION

AA952587

AA952587.1 GI:3115683

EST

Trypanosoma cruzi.

Trypanosoma cruzi

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma; Schizotrypanum.

1 (bases 1 to 267)

Verdun,R.E., Di Paolo,N.C., Urmenyi,T.P., Rondinelli,E., Frasch

,A.C.C. and Sanchez,D.O.

Gene discovery through expressed sequence tag sequencing in

trypanosoma cruzi

Infect. Immun. 66 (11), 5393-5398 (1998)

99003155

Contact: Sanchez D.O.

Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral

San Martin)

Av. Gral Paz entre Albarcellos y Constituyentes, INTI edificio 24

CP(1650) San Martin, Prov. de BS AS, Argentina

Tel: (54-1)752-9639 or (54-1)752-0021

Fax: (54-1)752-0021 or (54-1)752-9639

Email: dsanchez@inti.gov.ar

Seq primer: T7.

Location/Qualifiers

1..267

/organism="Trypanosoma cruzi"

/strain="Cl-Brenner"

/db_xref="taxon:5693"

/clone="1771"

/cell_type="epimastigote"

/note="cDNA library constructed with oligo dT primed

epimastigote mRNA and cloned in pT7318D phagemid with

modified polylinker (PHARMACIA)"

66 a 79 c 72 g 46 t 4 others

BASE COUNT

ORIGIN

Query Match 46.88; Score 22; DB 14; Length 267;
 Best Local Similarity 73.78; Pred. No. 11e+02;
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 7 aatgcgcccgcatttcgctgctgacgcccgaagctt 44

Db 200 AGTACTGCCGCAACTGCACCCGTCGCGCGCAAGCTT 237

RESULT 9

AI667900

LOCUS

DEFINITION

AI667900

AI667900.1 GI:4826272

EST

Trypanosoma

cruci cDNA clone n442.r 5', mRNA sequence.

AI667900

AI667900.1

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 826 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 253.

FEATURES
source

	M. FALCINA BOLLIGO.		
BASE COUNT	129 a	104 c	114 q
	94 t	94 t	6 others

Query Match 46.9%; Score 22; DB 1; Length 447

```
Query Match      46.8%; Score 22; DB 1; Length 447;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
```

RESULT 11
AT1975347

LOCUS
AI187524/C

DEFINITION

ACCESSION

VERSION

KEYWORDS
SOURCE

SOURCE ORGANISM

REFERENCE

AUTHORS

TITLE

1
2
3
4
5
6
7

JOURNAL
COMMENT

000000

/sex="male"
 /dev_stage="newly eclosed adults and pharate adults"
 /lab_host="Xl1 Blue MRF and SOLR"
 /note="Organ: antennae; Vector: Uni-ZAP XR; Site_1: EcoRI;
 Site_2: XhoI; The library was prepared by Stratagene using
 oligo-T priming and unidirectional cloning with an
 adaptor at the 5' end (GGCAGG) following the EcoRI
 site. The mRNA was prepared from antennae of late pupal
 and newly eclosed male moths. Clones were subcloned in
 vivo into mass into pBluescript maintained in SOLR cells
 for DNA sequencing."
 BASE COUNT 105 a 111 c 172 g 87 t
 ORIGIN

Query Match 46.8%; Score 22; DB 17; Length 475;
 Best Local Similarity 73.7%; Pred. No. 1.1e+02;
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 1 agctataatgcccgcacattatcgccgtgatggcgcg 38
 ||||| ||||| ||||| ||||| || |||||
 Db 142 AGTACGCTGCGCAGCATACTCGCGCGCATATGCGCG 105

RESULT 12
 BE353565 892 bp mRNA EST 09-AUG-2000
 LOCUS 601059673F1 NTH_MGC_10 Homo sapiens cDNA clone IMAGE:3445938 5',
 DEFINITION mRNA sequence.
 ACCESSION BE353565
 VERSION BE353565.1 GI:9764210
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 892)
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM8417 row: b column: 19
 High quality sequence stop: 612.

FEATURES
 source
 1..892
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3445938"
 /clone_lib="NIH_MGC_10"
 /cell_line="MGC36"
 /lab_host="DH10B"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
 Average insert size 1.5 Kb. Library prepared by Life
 Technologies."
 BASE COUNT 189 a 280 c 255 g 168 t
 ORIGIN

Query Match 46.8%; Score 22; DB 109; Length 892;
 Best Local Similarity 67.4%; Pred. No. 1.2e+02;
 Matches 31; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 QY 2 gctataatgcccgcacattatcgccgtgatggcgcgaagcttggg 47

Db 747 GCTCTAAGGGGCGCAGCATCTGGCGCGGTGATTGGA 792
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
 AV624476 436 bp mRNA EST 11-OCT-2000
 LOCUS AV624476 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
 DEFINITION reinhardtii cDNA clone LC078c03_r 5', mRNA sequence.
 ACCESSION AV624476
 VERSION AV624476.1 GI:10773653
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii

REFERENCE Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.

AUTHORS 1 (bases 1 to 436)
 Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
 Nakamura,Y. and Tabata,S.

TITLE Generation of Expressed Sequence Tags from low-CO2 and high-CO2
 adapted cells of Chlamydomonas reinhardtii
 JOURNAL DNA Res. (2000) In press
 COMMENT Contact: Erika Asamizu

The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
 Location/Qualifiers
 1..436
 /organism="Chlamydomonas reinhardtii"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone="LC078c03_r"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI; The cDNA library was constructed from cells cultured
 in a carbon stress acclimatized condition in which carbon
 dioxide concentration in the bubbling gas was changed from
 5% to 0.04%
 BASE COUNT 69 a 157 c 116 g 94 t
 ORIGIN

Query Match 46.4%; Score 21.8; DB 37; Length 436;
 Best Local Similarity 78.8%; Pred. No. 1.4e+02;
 Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 11 cggcgccattatcgccgtgatcgccgcaagct 43
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 402 CGGCCGCGCATTCGCATGATGCGCGCCGCGCT 434

RESULT 14
 AZ342872/c 558 bp DNA GSS 29-SEP-2000
 LOCUS AZ342872 Mouse 10kb plasmid UUC1M library Mus musculus genomic
 DEFINITION clone UUC1M007601 F, DNA sequence.
 ACCESSION AZ342872
 VERSION AZ342872.1 GI:10420544
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 558)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 JOURNAL plasmid inserts
 COMMENT Mouse whole genome scaffolding with paired end reads from 10kb
 Unpublished (2000)
 Contact: Robert B. Weiss

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 12:07:04 ; Search time 96.53 Seconds
(without alignments)
182.908 Million cell updates/sec

Title: US-09-101-423B-8

Perfect score: 47

Sequence: 1 agctaatgcgcgcata.....tgatcgccgaagcttga 47

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36.*

1:	/cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2:	/cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3:	/cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5:	/cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6:	/cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8:	/cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9:	/cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11:	/cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12:	/cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14:	/cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15:	/cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17:	/cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18:	/cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20:	/cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21:	/cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	43	91.5	45	18 T72784	DNA tag. Syntheti
C 2	21.8	46.4	1503	19 V15518	Hamster oral papil
C 3	21.8	46.4	7647	19 V15519	Hamster oral papil
C 4	21.4	45.5	42	18 T61221	Primer HUCKN.NOT f
5	21.2	45.1	500	20 X91235	T. gondii immunoge
6	21.2	45.1	1818	21 Z46842	Semaphorin K1 gene
7	21.2	45.1	3481	21 Z21501	DNA sequence of so
C 8	20.8	44.3	8907	21 Z36926	Nucleotide sequenc
C 9	20.8	44.3	9181	20 X84323	Stealth virus nucl
10	20.6	43.8	144460	21 Z93815	Olfactory receptor
11	20.4	43.4	41	15 O70114	Lambda INNER prime
12	20.4	43.4	4260	9 N81768	Sequence encoding

13	20.4	43.4	4380	9 N80222	Sequence of Mycoba
14	20.4	43.4	4380	19 V05708	Mycobacterium tube
C 15	20.2	43.0	2885	20 X40055	Colon cancer assoc
16	20	42.6	2008	21 A09260	Human alpha-2-delt
17	20	42.6	3203	11 Q06760	Sequence of beta-g
18	20	42.6	3598	21 A09261	Human alpha-2-delt
19	20	42.6	3770	21 A09253	Human alpha-2-delt
20	20	42.6	5059	21 Z36227	DNA encoding a per
C 21	19.8	42.1	1173	20 X91764	Porphyromonas ging
C 22	19.8	42.1	1218	20 X91630	Porphyromonas ging
23	19.8	42.1	2178	20 Z12261	Neisseria gonorrhoe
24	19.8	42.1	3415	17 T36481	Human integrin bet
25	19.8	42.1	6727	18 T88014	Murine IL-5 cDNA g
26	19.8	42.1	7673	19 V58229	Omega-cyclodextrane
27	19.6	41.7	3183	20 X26611	Trypanosoma cruzi
C 28	19.4	41.3	939	21 Z29251	Rifin 3193 gene.
C 29	19.4	41.3	2943	17 T16480	SAB virus gB glyco
C 30	19.2	40.9	619	21 Z53637	Neisseria gonorrhoe
C 31	19.2	40.9	834	21 Z53639	Neisseria meningit
C 32	19.2	40.9	1842	21 Z56349	Escherichia coli f
33	19.2	40.9	1830121	17 T42063	Haemophilus influe
C 34	19	40.4	638	18 T45880	Human colon specif
C 35	19	40.4	638	19 V16668	Polynucleotide seq
36	19	40.4	924	21 A12829	Pseudomonas cepaci
37	19	40.4	3087	11 Q03634	Sequence for the 3
C 38	19	40.4	4394	13 Q21604	Alpha galactosidas
39	19	40.4	5474	16 Q90251	Tuberosus sclerosi
C 40	18.8	40.0	1095	21 Z54498	Neisseria meningit
41	18.8	40.0	1229	20 Z42169	Human normal bladd
C 42	18.8	40.0	1797	20 X83965	Salmonella typhimu
43	18.8	40.0	2302	21 A15621	Human heat shock p
44	18.8	40.0	2302	21 A15622	Human heat shock p
45	18.8	40.0	2465	18 T58086	Human heat shock p

ALIGNMENTS

RESULT 1

T72784/c

ID T72784 standard; DNA; 45 BP.

XX T72784;

XX AC

XX DT 22-SEP-1997 (first entry)

XX DE DNA tag.

XX KW Metastasis-inducing DNA; Met-DNA; cancer; diagnosis; ds.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT misc_feature 1..2

FT FT /*tag= a

FT FT /*note= "5' single stranded overhang"

FT FT misc_feature 45

FT FT /*tag= b

FT FT /*note= "5' overhang on complementary strand of 4 bases with sequence 5'-AGCT-3'"

XX WO9725443-A1.

XX PD 17-JUL-1997.

XX PF 10-JAN-1997; 97WO-GB000074.

XX PR 10-JAN-1996; 96GB-0000470.

XX PA (UYLI-) UNIV LIVERPOOL.

XX PI Barracough BR, Rudland PS;

XX XX

DR WPI; 1997-372878/34.
 XX New isolated metastasis-inducing DNA - used to develop products to
 PT identify and treat patients at risk from metastatic tumours
 XX
 XX Claim 6; Page 24; 38pp; English.
 XX
 CC A DNA tag (T72784) having HindIII, SfiI, NotI and a defective
 CC HindIII site is used to tag human DNA from malignant, metastatic
 CC cancer cells in a novel method of screening and recovering
 CC metastasis-inducing DNA (Met-DNA). This method involves
 CC transferring the human DNA into a cell line (pref. rat Rama 37)
 CC that produces only benign, non-metastasising tumours when injected
 CC into a syngeneic animal, infecting those animals having metastasising
 CC tumours, and recovering the Met-DNA (see also T72785-90) from them.
 CC The human DNA is tagged to assist in its removal or insertion from
 CC or into a host or vector. The tagging procedure overcomes the
 CC problem of identifying the inserted human DNA sequences in the rat
 CC genome of the transfected cells.
 XX
 SQ Sequence 45 BP; 11 A; 13 C; 12 G; 9 T; 0 other;

Query Match 91.5%; Score 43; DB 18; Length 45;
 Best Local Similarity 100.0%; Pred. No. 4e-08;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ataatcgccgcgcatttcgctgctgacgcgcgcgaagcttga 47
 |||||
 Db 45 ATATGCGCGCGCATATTCGCTGATCGCGCGCAAGCTTGA 3

RESULT 2

VI5518/c
 ID VI5518 standard; DNA; 1503 BP.

XX
 AC VI5518;

DT 22-MAY-1998 (first entry)

XX Hamster oral papilloma virus L1 DNA.

XX L1 DNA; L1 protein; diagnosis; papilloma virus infectious disease;
 KW cancer; screening; antitumour agent; antisense treatment; ss.

XX Hamster oral papilloma virus.

OS
 XX JP10042875-A.

PD 17-FEB-1998.

XX 06-AUG-1996; 96JP-0207143.

XX 06-AUG-1996; 96JP-0207143.

XX (TORA) TORAY IND INC.

XX WPI; 1998-186869/17.

DR P-PSDB; W47224.

XX DNA encoding hamster oral papilloma virus protein - useful for, e.g.
 PT diagnosing cancer related papilloma virus infections

XX Claim 10; Pages 13-14; 17pp; Japanese.

XX The present sequence is the hamster oral papilloma virus L1 DNA.

XX The L1 protein and DNA can be used for the diagnosis of papilloma

CC virus infectious diseases related to cancer, screening for

CC antitumour agents and antisense treatment.

XX Sequence 1503 BP; 434 A; 346 C; 337 G; 386 T; 0 other;

Query Match 46.4%; Score 21.8; DB 19; Length 1503;
 Best Local Similarity 70.7%; Pred. No. 9.9;
 Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 tataatcgccgcgcatttcgctgctgacgcgcgcgaagctt 44
 |||||
 Db 866 TATGATGCATCAGCCTTTTGATCCTGATCAGCGCAGGTTT 826

RESULT 3

VI5519/c
 ID VI5519 standard; DNA; 7647 BP.

XX
 AC VI5519;

XX 22-MAY-1998 (first entry)

XX Hamster oral papilloma virus genomic DNA.

XX Diagnosis; papilloma virus infectious disease;
 KW cancer; screening; antitumour agent; antisense treatment; ss.

XX Hamster oral papilloma virus.

XX JP10042875-A.

PD 17-FEB-1998.

XX 06-AUG-1996; 96JP-0207143.

XX 06-AUG-1996; 96JP-0207143.

XX (TORA) TORAY IND INC.

XX WPI; 1998-186869/17.

XX DNA encoding hamster oral papilloma virus protein - useful for, e.g.
 PT diagnosing cancer related papilloma virus infections

XX Claim 11; Pages 14-16; 17pp; Japanese.

XX The present sequence is a hamster oral papilloma virus genomic
 CC DNA. The DNA can be used for the diagnosis of papilloma virus
 CC infectious diseases related to cancer, screening for antitumour
 CC agents and antisense treatment.

XX Sequence 7647 BP; 2102 A; 1703 C; 1894 G; 1948 T; 0 other;

Query Match

46.4%; Score 21.8; DB 19; Length 7647;
 Best Local Similarity 70.7%; Pred. No. 12;
 Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 4 tataatcgccgcgcatttcgctgctgacgcgcgcgaagctt 44
 |||||
 Db 6514 TATGATGCATCAGCCTTTTGATCCTGATCAGCGCAGGTTT 6474

RESULT 4

T61221/c
 ID T61221 standard; DNA; 42 BP.

XX
 AC T61221;

XX 17-OCT-1997 (first entry)

XX Primer HUCKN.NOT for light chain constant region cDNA.

XX Primer; polymerase chain reaction; PCR; amplification; light chain;
 KW immunoglobulin; Ig; constant; region; production;
 KW recombinant; antibody; B cell; diagnosis; therapy; ss.

(HOFF) HOFFMANN LA ROCHE & CO AG F.

Hosino T, Miyazaki T, Ojima S, Shinjoh M, Tomiyama N;
WPI; 1999-134646/12.
P-PSDB; W95019, W95020.

New D-sorbitol dehydrogenase gene and recombinant protein - useful
for production of L-sorbose, an intermediate in vitamin C production

Claim 4; Fig 3A-D; 39pp; English.

This represents the DNA sequence of the SLDH gene encoding a protein with
sorbitol dehydrogenase (SLDH) activity. The DNA also encodes an open
reading frame (ORF2) product upstream of the SLDH open reading frame,
comprising for SLDH activity in vivo. Host cells transformed by a vector
containing the SLDH DNA sequence are used for the recombinant expression
of the sorbitol dehydrogenase. Recombinant D-sorbitol dehydrogenase is
useful for producing L-sorbose from D-sorbitol. L-sorbose is an important
intermediate in vitamin C production.

Sequence 3481 BP; 674 A; 1018 C; 1001 G; 788 T; 0 other;

Query Match 45.1%; Score 21.2; DB 20; Length 3481;
Best Local Similarity 76.5%; Pred. No. 19;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 13 gcgcgcatattggcctgatcgccgcgaagcttgg 46
|| || || || || || || || || || || || || || || ||
Db 355 gcgcggtattcatgtctatggccgcgacgcttgg 388

RESULT 8
Z36926/c
ID Z36926 standard; DNA; 8907 BP.
XX AC Z36926;
XX XX
XX XX
DT 13-MAR-2000 (first entry)
XX XX
DE Nucleotide sequence of the genome of Stealth virus clone 313.
XX XX
KW Prototype Stealth virus clone; atypically structured virus;
KW vacuulating cytopathic effect; cytopathic virus; virus detection;
KW malignancy; multiple myeloma; lymphoma; brain tumour; breast cancer;
KW salivary gland tumour; Alzheimer's disease; Parkinson's disease;
KW spongiform encephalopathy; multiple sclerosis; schizophrenia;
KW manic depression; major depression; personality disorder; autism;
KW Rett's syndrome; attention deficit; oppositional defiance; aggression;
KW anorexia nervosa; bulimia; multi-system illness; virus infection; ss.
XX XX
OS Stealth virus.
XX XX
PN WC9960101-A1.
XX XX
PD 25-NOV-1999.
XX XX
PF 19-MAY-1999; 99WO-US11185.
XX XX
PR 19-MAY-1998; 98US-0081708.
XX XX
PA (MARTIN/) MARTIN W J.
XX XX
PI Martin WJ;
XX XX
WI WPI; 2000-072436/06.

Isolated viruses, used to develop products for detection, prevention
and treatment of stealth virus infections _
Disclosure; Page 34-37; 50pp; English.

PT New nucleic acid encoding enterokinase activity - and related
PT vectors, host cells, expression products and antibodies are
PT useful in treating digestive disorders and for cleaving fusion
PT proteins
XX
PS Example 6; Page 19; 50pp; English.
XX
CC Nested oligonucleotide primers were synthesised which were
CC complementary to the lambda DNA sequence adjacent to the cloning site
CC for the cDNA insertions (see Q70113-14). In addition, primers were
CC designed which were complementary to the plus strand of the most 5'
CC region of the EK coding sequence (see Q70115-16). Subcloning and
CC sequencing provided a 116 amino acid EK heavy chain. This protein
CC contained regions corresp. to tryptic peptides (see R57295-9).
XX
SQ Sequence 41 BP; 12 A; 10 C; 11 G; 8 T; 0 other;

Query Match 43.4%; Score 20.4; DB 15; Length 41;
Best Local Similarity 71.1%; Pred. No. 22;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 5 ataatgcccgcattatcgccctgacgcgcgcgaagc 42
| ||||| ||||| ||||| |||||
Db 4 agaatgcccgcgaagtctcagcctggttaagtccaagc 41

RESULT 12
N81768
ID N81768 standard; DNA; 4260 BP.
XX
AC N81768;
XX
XX 29-DEC-1990 (first entry)
DT
DE Sequence encoding Mycobacterium tuberculosis 540 and 517 AA residue
DE proteins.
XX
KW Diagnosis; assay; M.bovis; vaccine; ds.
XX
OS Mycobacterium tuberculosis.
XX
XX
FH Key Location/Qualifiers
FT CDS 252..1874
FT FT /*tag= a
FT FT /label=540 AA protein
FT FT /note="P81351"
FT CDS complement (3948..2395)
FT FT /*tag= b
FT FT /label=517 AA protein
FT FT /note="P81868"
XX
XX WO8806591-A.
XX
XX
XX 07-SEP-1988.
XX
XX 25-FEB-1988; 88WO-US000598.
XX
XX 24-FEB-1988; 88US-0159667.
XX 06-FEB-1987; 87US-0019529.
XX
XX (SCRI-) SCRIPPS CLINIC & RE.
XX
XX Shinnick T, Houghten R;
XX
XX WPI: 1988-271136/38.
XX P-PSDB; P81351, P81868.
XX Recombinant mycobacterial peptide(s) -
PT used in assays for diagnosis of infection, for producing
PT vaccines and for producing antibodies
XX
PS Disclosure; Fig 2a-2d; 116pp; English.

XX An isolated DNA molecule that consists essentially of the nucleotide
CC sequence that corresponds to the sequence represented by position 3950 to
CC about 2390 and from position 3948 through position 2398 of N81768 is
CC claimed. Also claimed is a peptide sequence that consists of a 5-40 AA
CC residue sequence that corresponds to a sequence of the 540 AA residue
CC protein (P81351) or the 517 AA residue protein (P81868) coded for by the
CC DNA sequence. The proteins can be used for determining previous
CC immunological exposure of a mammal to M.tuberculosis or M.bovis and
CC for producing a vaccine.
XX
SQ Sequence 4260 BP; 733 A; 1332 C; 1481 G; 714 T; 0 other;

Query Match 43.4%; Score 20.4; DB 9; Length 4260;
Best Local Similarity 80.0%; Pred. No. 39;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 9 tgcggccgcattatcgccctgacgcgcgcgc 38
| ||||| ||||| ||||| |||||
Db 3481 tgcggccgcctgttcgcgcctgctgctgcgc 3510

RESULT 13
N80222
ID N80222 standard; DNA; 4380 BP.
XX
AC N80222;
XX
XX 19-MAR-1991 (first entry)
DT
DE Sequence of Mycobacterium tuberculosis DNA contg. gene encoding 65
DE protein.
XX
KW Antigen; vaccine; ds.
XX
OS Mycobacterium tuberculosis.
XX
XX
FH Key Location/Qualifiers
FT CDS 192..1874
FT FT /*tag= a
FT CDS complement (2398..4101)
FT FT /*tag= b
XX
XX WO8805823-A.
XX
XX 11-AUG-1988.
XX
XX 01-FEB-1988; 88WO-US000281.
XX
XX 02-FEB-1987; 87US-0010007.
XX
XX (WHIT-) WHITEHEAD INST BIOM.
XX
XX Husson RN, Young RA, Shinnick TM;
XX
XX WPI: 1988-235175/33.
XX P-PSDB; P80215, P80216
XX
XX Genes encoding Mycobacterium tuberculosis protein antigens -
PT useful for developing reagents for diagnosis, prevention and
PT treatment of tuberculosis
XX
XX Claim 12; Fig 8; 82pp; English.
XX
XX The gene was isolated by probing a lambda gt11 expression library of
CC M.tuberculosis DNA with monoclonal antibodies directed against
CC M.tuberculosis-specific antigens. The 19KD, 71KD and the 65KD proteins
CC and genes are claimed, and so is a vaccine comprising DNA encoding
CC M.tuberculosis protein in a recombinant vaccine vector. P80216 is
CC encoded on the complementary strand.
XX
XX Sequence 4380 BP; 757 A; 1373 C; 1512 G; 738 T; 0 other;

Query Match 43.4%; Score 20.4; DB 9; Length 4380;
 Best Local Similarity 80.0%; Pred. No. 40;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 9 tgcggcgcatattcgccgtgctgcgcgc 38
 ||||| ||| ||||| |||||
 Db 3601 tgcggcgcatattcgccgtgctgcgcgc 3630

RESULT 14

V05708 ID V05708 standard; DNA; 4380 BP.

XX AC V05708;

XX DT 22-JUN-1998 (first entry)

XX DE Mycobacterium tuberculosis 65 kDa heat shock protein gene.

XX KW Heat shock protein; Mt Hsp65; autoimmune disease; immunotherapy;
 gene therapy; rheumatoid arthritis; multiple sclerosis; ds.

XX OS Mycobacterium tuberculosis.

XX FH Key Location/Qualifiers
 XX CDS 252..1874

XX FT /*tag= a
 XX FT /product= 65 kDa heat shock protein

XX PN W09746253-A2.

XX PD 11-DEC-1997.

XX PF 03-JUN-1997; 97WO-US09427.

XX PR 03-JUN-1997; 97US-0019100.

XX PR 03-JUN-1996; 96US-0019100.

XX PA (AURA-) AURAGEN INC.

XX PI Haynes JR, Prayaga SK, Ramshaw TA;

XX PI WPI; 1998-041892/04.

XX DR P-PSDB; W44702.

XX PT Treatment of autoimmune diseases - by administering
 XX PT autoantigen-coated particles or autoantigen-encoding nucleic acid
 XX PT construct

XX PS Example 2; Page 55-59; 72pp; English.

XX CC This DNA sequence encodes the 65 kDa heat shock protein (see
 CC W44702), designated Mt Hsp65, of Mycobacterium tuberculosis. This
 CC protein cross-reacts with a component of articular cartilage, human
 CC Hsp60, that is up-regulated in the joints of arthritic patients. A
 CC claimed method for treating or preventing an autoimmune disease in
 CC a mammal comprises: (a) providing a particle coated with an antigen
 CC against which an immune response is mounted in the autoimmune
 CC disease; (b) delivering the particle into the recipient cell of the
 CC mammal; and (c) repeating step (b) until either a reduction in a
 CC cytotoxic immune response or a desensitizing immune response is
 CC induced in the mammal. Alternatively, step (a) comprises providing
 CC a nucleic acid construct comprising a coding sequence for the
 CC antigen, operably linked to control elements such that the coding
 CC sequence can be transcribed and translated in a recipient cell, and
 CC delivering the construct to the recipient cell using a gene gun.
 CC The antigen of step (a) is selected from collagen, Mt Hsp65,
 CC myelin basic protein, myelin oligodendrocyte glycoprotein,
 CC proteolipid protein, and epitopes thereof. These antigens mitigate
 CC cytotoxic responses and elicit antigen desensitisation. The method
 CC is used especially for treating rheumatoid arthritis or multiple

CC sclerosis. It represents a novel use for the known Mt Hsp65 gene.
 XX Sequence 4380 BP; 757 A; 1371 C; 1514 G; 738 T; 0 other;
 SQ

Query Match 43.4%; Score 20.4; DB 19; Length 4380;
 Best Local Similarity 80.0%; Pred. No. 40;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 9 tgcggcgcatattcgccgtgctgcgcgc 38
 ||||| ||| ||||| |||||
 Db 3601 tgcggcgcatattcgccgtgctgcgcgc 3630

RESULT 15

X40055/c

XX ID X40055 standard; DNA; 2885 BP.

XX AC X40055;

XX DT 02-JUL-1999 (first entry)

XX DE Colon cancer associated gene.

XX KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer; ss.

XX OS Homo sapiens.

XX XX W09904265-A2.

XX PN 28-JAN-1999.

XX PD 15-JUL-1998; 98WO-US14679.

XX PR 22-JUN-1998; 98US-0102322.

XX PR 17-JUL-1997; 97US-0896164.

XX PR 10-OCT-1997; 97US-0061599.

XX PR 10-OCT-1997; 97US-0061765.

XX PR 10-OCT-1997; 97US-0948705.

XX PR 11-OCT-1997; 97GB-0021697.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

XX PI Pfeundschoh M, Sahin U, Scanlan MJ, Stockert E;

XX PI Tureci O;

XX DR WPI; 1999-132448/11.

XX PT New isolated cancer associated nucleic acids and polypeptides -
 XX PT isolated using sera from cancer patients, used to develop products
 XX PT for the diagnosis, monitoring or treatment of cancers

XX PS Claim 67; Page 651-652; 787pp; English.

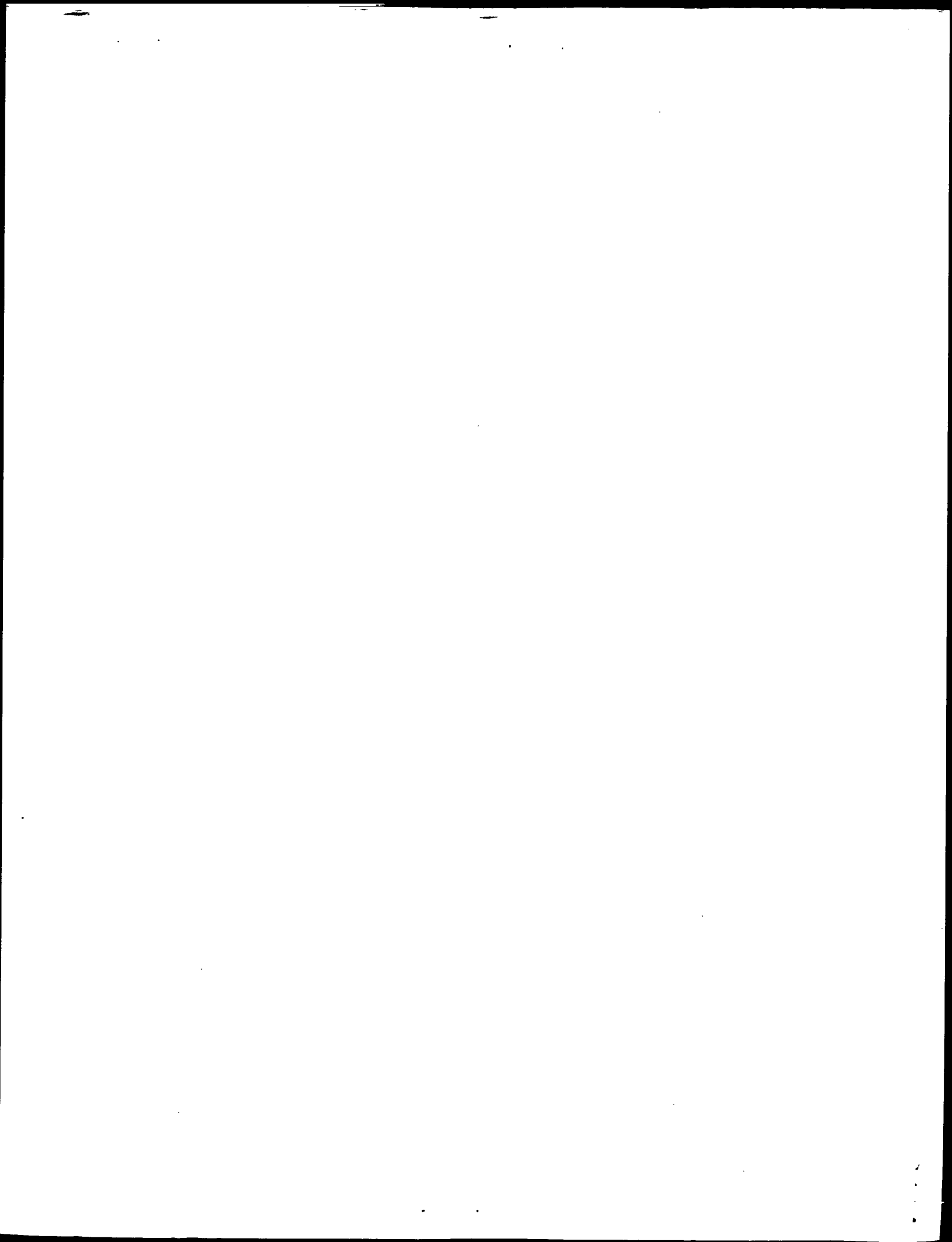
XX CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.

XX SQ Sequence 2885 BP; 626 A; 901 C; 838 G; 520 T; 0 other;

Query Match 43.0%; Score 20.2; DB 20; Length 2885;
Best Local Similarity 75.8%; Pred. No. 45;
Matches 25; Conservative 0; Mismatches 8; Indels 0;

y 10 ggggccgcataattcggcctgatcggccgcaagc 42
||| ||| ||| ||| ||| ||| ||| ||| |||
b 2747 GCTGCCGACGTGGGCCTGTTGGCCCCCACC 27

Search completed: March 4, 2001, 12:07:09
Job time: 5664 sec



; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Melnert, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-200-900A-26

Query Match 43.4%; Score 20.4; DB 1; Length 41;
Best Local Similarity 71.1%; Pred. No. 11;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 5 ataatgcgcgcgatattcgcctgctgcgcgcgaagc 42
| | | | | | | | | | | | | | | | | | | | | |
Db 4 AGAATGCGCGCAAGTTACGCTGGTTAAGTCCAAGC 41

RESULT 3
PCT-US94-00616-26
; Sequence 26, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00616
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US94-00616-26

Query Match 43.4%; Score 20.4; DB 4; Length 41;
Best Local Similarity 71.1%; Pred. No. 11;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 5 ataatgcgcgcgatattcgcctgctgcgcgcgaagc 42
| | | | | | | | | | | | | | | | | | | | | |
Db 4 AGAATGCGCGCAAGTTACGCTGGTTAAGTCCAAGC 41

RESULT 4
US-08-948-705-2/c
; Sequence 2, Application US/08948705A
; Patent No. 6043084
; GENERAL INFORMATION:

; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND METHODS FOR DIAGNOSING AND
; TITLE OF INVENTION: TREATING COLON CANCER
; FILE REFERENCE: LUD-5506-JEL/NDH
; CURRENT APPLICATION NUMBER: US/08/948,705A
; CURRENT FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2885
; TYPE: DNA
; ORGANISM: Homo Sapiens
; US-08-948-705-2

Query Match 43.0%; Score 20.2; DB 3; Length 2885;
Best Local Similarity 75.8%; Pred. No. 24;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 10 gcgcgccgcgatattcgcctgctgcgcgcgaagc 42
| | | | | | | | | | | | | | | | | | | | | |
Db 2747 GCTGGCGGCGAGCTGGCCGTGTCGGCCCCCACC 2715

RESULT 5
US-08-054-077C-1
; Sequence 1, Application US/08054077C
; Patent No. 5527679
; GENERAL INFORMATION:
; APPLICANT: HEMLER, MARTIN E.
; APPLICANT: RAMASWAMY, HEMAVATHI
; TITLE OF INVENTION: HUMAN INTEGRIN 5 SUBUNIT PROTEIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/054,077C
; FILING DATE: 27-APR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694314
; FILING DATE: 01-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 40937
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3415 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: mat peptide


```

1  /  ADDRESSEE:  Intellectual Property Services
2  /
3  /  ADDRESSEE:  Battelle Memorial Institute
4  /
5  /  ADDRESSEE:  PNNL P.O. Box 999
6  /
7  /  STREET:  Washington Way
8  /
9  /  CITY:  Richland
10 /
11 /  STATE:  Washington
12 /
13 /  COUNTRY:  U.S.A.
14 /
15 /  ZIP:  99352
16 /
17 /  COMPUTER READABLE FORM:
18 /
19 /  MEDIUM TYPE:  Diskette 3.50", 1.44 Mb storage
20 /
21 /  COMPUTER:  IBM PC/XT/AT
22 /
23 /  OPERATING SYSTEM:  MS-DOS
24 /
25 /  SOFTWARE:  Word Processor (WordPerfect 5.1)
26 /
27 /  CURRENT APPLICATION DATA:
28 /
29 /  APPLICATION NUMBER:  US/08/853,659A
30 /
31 /  FILING DATE:  Unknown
32 /
33 /  CLASSIFICATION:  435
34 /
35 /  PRIOR APPLICATION DATA:
36 /
37 /  APPLICATION NUMBER:  none
38 /
39 /  FILING DATE:  n/a
40 /
41 /  INFORMATION FOR SEQ ID NO: 28:
42 /
43 /  SEQUENCE CHARACTERISTICS:
44 /
45 /  LENGTH:  1797 bases
46 /
47 /  TYPE:  nucleotide
48 /
49 /  STRANDEDNESS:  double stranded
50 /
51 /  TOPOLOGY:  linear
52 /
53 /  FEATURE:
54 /
55 /  OTHER INFORMATION:  the coding nucleotides of
56 /
57 /  OTHER INFORMATION:  NO:28 correspond to nucle
58 /
59 /  OTHER INFORMATION:  24426 of SEQ ID NO:2
60 /
61 /  US-08-853-659A-28

```

Query Match	40.0%;	Score 18.8;	DB 2;
Best Local Similarity	68.4%;	0; Pred. Mismatches	12;
Matches	26;	Conservative	0;
QY	3	ctataatgcggccgcatattgcggcgctgacgcgcgcga	40
Db	1101	CGTGAATCGCGGTGTTTTTCGCGTATCTTCATCAA	1064

```

RESULT 12
US-08-853-659A-6/c
; Sequence 6, Application US/08853659A
; Patent No. 5925522
; GENERAL INFORMATION:
; APPLICANT: Wong, K. K.; Saffer, J. D.
; TITLE OF INVENTION: A Salmonella Sequence, Methods Of Detection
; TITLE OF INVENTION: Of A
; TITLE OF INVENTION: Salmonella Sequence, And Methods Of Detection Of Salmonella
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Paul W. Zimmerman
; ADDRESSEE: Intellectual Property Services
; ADDRESSEE: Battelle Memorial Institute
; ADDRESSEE: PNNL P.O. Box 999
; STREET: Washington Way
; CITY: Richland
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 99352
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.50", 1.44 Mb storage
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (WordPerfect 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/853,659A
; FILING DATE: Unknown
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: none

```

```

; FILING DATE: n/a
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8967 bases
; TYPE: nucleotide
; STRANDEDNESS: double stranded
; TOPOLOGY: linear
; FEATURE:
; OTHER INFORMATION: SEQ ID NO:6 corresponds to
; OTHER INFORMATION: nucleotides 15735 through 24701 of SEQ ID NO:2
US-08-853-659A-6

```

Query Match	40.0%;	Score 18.8;	DB 2;	Length 8967;
Best Local Similarity	68.4%;	Pred. No. 98;		
Matches 26;	Conservative 0;	Mismatches 12;	Indels 0;	Gaps 0;
QY	3	ctataatgcccgcgcataattcggcctgatcgccgcgcaa	40	
Db	7996	CTGTAATGCCGGTGTTTTTCTGCTGATCTTCATCAA	7959	

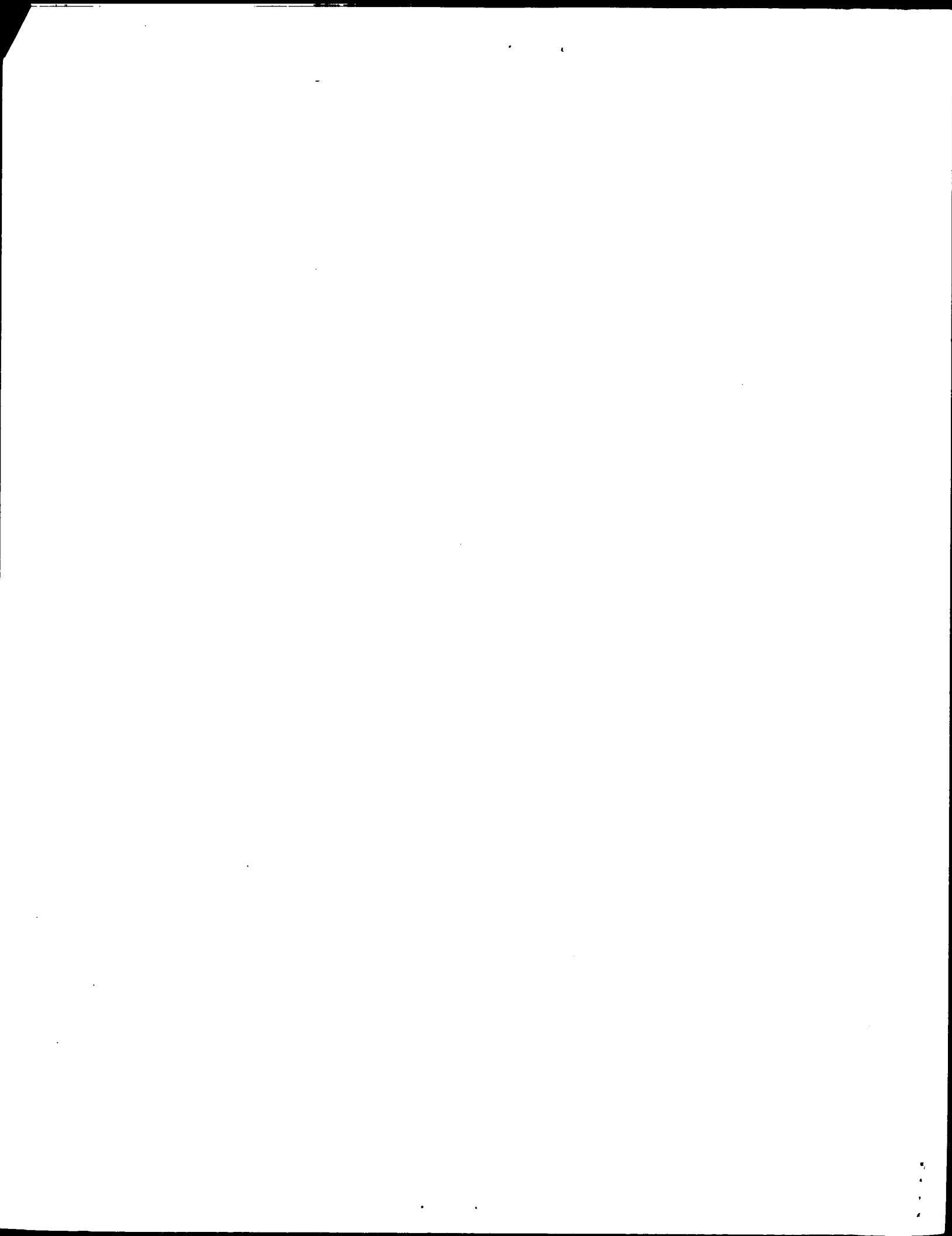
```

RESULT 13
US-08-853-659A-9
; Sequence 9, Application US/08853659A
; Patent No. 592552
; GENERAL INFORMATION:
; APPLICANT: Wong, K.K.; Saffer, J.D.
; TITLE OF INVENTION: A Salmonella Sequence, Methods Of Detection
; TITLE OF INVENTION: Of A
; TITLE OF INVENTION: Salmonella Sequence, And Methods Of Detection Of Salmonella
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Paul W. Zimmerman
; ADDRESSEE: Intellectual Property Services
; ADDRESSEE: Battelle Memorial Institute
; ADDRESSEE: PNNL P.O. Box 999
; STREET: Washington Way
; CITY: Richland
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 99352
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.50", 1.44 Mb storage
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (WordPerfect 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/853,659A
; FILING DATE: Unknown
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: none
; FILING DATE: n/a
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8967 bases
; TYPE: nucleotide
; STRANDEDNESS: double stranded
; TOPOLOGY: linear
; US-08-853-659A-9

```

Query Match	40.0%	Score 18.8:	DB 2:	Length 8967;
Best Local Similarity	68.4%;	Pred. No. 98;	-	-
Matches 26;	Conservative 0;	Mismatches 12;	Indels 0;	Gaps 0;
Qy	3	c t a t a a t c g c g c g c a t a t t c g c c t g a t c g c g c g c a a	40	
Dh	972	C T G T A A T C S C C G G T G T T T T T T T C G C T A C T T C A P T C A A	1009	

RESULT 14



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 12:03:42 ; Search time 1021.9 Seconds
(without alignments)
235.379 Million cell updates/sec

Title: US-09-101-423B-8
Perfect score: 47
Sequence: 1 agctataatgcggccgcata.....tgatcgccgcgaagcttgga 47

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues 2236266
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_ph: *
6: gb_pl1: *
7: gb_pl2: *
8: gb_pri: *
9: gb_pr2: *
10: gb_pr3: *
11: gb_ro: *
12: gb_sy: *
13: gb_un: *
14: em_fun: *
15: em_hum1: *
16: em_hum2: *
17: em_in: *
18: em_om: *
19: em_or: *
20: em_ov: *
21: em_pat: *
22: em_ph: *
23: em_pl: *
24: em_ro: *
25: em_sts: *
26: em_sy: *
27: em_un: *
28: em_vi: *
29: gb_ba3: *
30: gb_in1: *
31: gb_in2: *
32: gb_in3: *
33: gb_pl3: *
34: gb_pr4: *
35: em_ba1: *
36: em_ba2: *
37: em_htg1: *
38: em_htg2: *
39: em_htg3: *
40: em_htg4: *
41: em_htg5: *
42: em_htg6: *
43: em_htg7: *

44: em_htg8: *
45: em_htg9: *
46: em_htg10: *
47: em_hum3: *
48: em_hum4: *
49: em_hum5: *
50: em_hum6: *
51: gb_pr5: *
52: gb_pr6: *
53: gb_pr7: *
54: gb_htg1: *
55: gb_htg2: *
56: gb_htg3: *
57: gb_htg4: *
58: gb_htg5: *
59: gb_htg6: *
60: gb_htg7: *
61: gb_htg8: *
62: gb_htg9: *
63: gb_htg10: *
64: gb_htg11: *
65: gb_htg12: *
66: gb_htg13: *
67: gb_htg14: *
68: gb_htg15: *
69: gb_htg16: *
70: gb_htg17: *
71: gb_htg18: *
72: gb_htg19: *
73: gb_htg20: *
74: gb_htg21: *
75: gb_htg22: *
76: gb_htg23: *
77: gb_sts1: *
78: gb_sts2: *
79: gb_vil1: *
80: gb_vil2: *
81: gb_pat1: *
82: gb_pat2: *
83: em_htg0: *
84: gb_htg24: *
85: gb_pr8: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID.	Description
C 1	24.8	52.8	86896	2	RCU57682	U57682 Rhodobacter
C 2	23.4	49.8	25458	29	SCE22	AL355832 Streptomy
C 3	23.4	49.8	177241	52	HS402G11	AL022328 Human DNA
C 4	23	48.9	2042	11	MUSGATGABA	L32178 Mus cookii
C 5	23	48.9	4054	11	RATGABAT	M59742 Rat GABA tr
C 6	23	48.9	4074	11	MUSGABAX	M52378 Mus musculu
C 7	23	48.9	4402	11	MUSCOOK	M97512 Mus cookii
C 8	23	48.9	10908	1	AE002294	AE002294 Chlamydia
C 9	22.6	48.1	2013	2	BACCCOTJABC	L38014 Bacillus su
C 10	22.6	48.1	17103	59	AC017738	AC017738 Drosophil
C 11	22.6	48.1	18622	1	AF012532	AF012532 Bacillus
C 12	22.6	48.1	101962	55	AC009915	AC009915 Drosophil
C 13	22.6	48.1	213190	2	BSUB0004	299107 Bacillus su
C 14	22.6	48.1	226159	30	AE003738	AE003738 Drosophil
C 15	22.2	47.2	1950	2	AVIGLNA	M57275 A.vinelandi
C 16	22.2	47.2	2565	1	AF030293	AF030293 Rathayiba
C 17	22.2	47.2	2565	2	RRAJ2069	AJ002069 Rathayiba
C 18	22.2	47.2	6254	32	DMU13637	U13637 Drosophila
C 19	22.2	47.2	14724	58	AC015103	AC015103 Drosophil
C 20	22.2	47.2	151673	62	AC022365	AC022365 Homo sapi
C 21	22.2	47.2	218971	56	AC011697	AC011697 Drosophil

AUTHORS	Liu,Q.-R., Mandiyan,S., Nelson,H. and Nelson,N.
TITLE	A family of genes encoding neurotransmitter transporters
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. (1992) In press
FEATURES	Location/Qualifiers
source	1. 4402
	/organism="Mus cookii"
	/db_xref="taxon:10098"
BASE COUNT	1055 a 1090 c 1230 g 1027 t
ORIGIN	

```

Query Match      48.9%; Score 23; DB 11; Length 4402;
Best Local Similarity 68.1%;
Pred. No. 3.3e+02;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 agctataatgcggccgcataatgcctgatcggccgcgaagcttggga 47
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3973 AGAATATCTGCGCCCTGAGAAATGSCCTGAGCAGCCGACGGCTGCA 4019

```

RESULT	8
LOCUS	AEO02294/c
DEFINITION	AEO02294
ACCESSION	Chlamydia muridarum, section 26 of 85 of the complete genome.
VERSION	AEO02294.AE002160
KEYWORDS	AEO02294.1 GI:7190305
SOURCE	.
ORGANISM	Chlamydia muridarum. Chlamydia muridarum Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. 1 (bases 1 to 10908) Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.
REFERENCE	
AUTHORS	

JOURNAL	Nucleic Acids Res.	28	(6)	1397-1406	(2000)
MEDLINE	20150255				
INDEXED	10684935				
REFERENCE	2	(bases 1 to 10908)			
AUTHORS	Read, T.D., Brunham, R., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Winn, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-MAR-2000)	The Institute for Genomic Research,			9712
	Medical Center Dr,	Rockville, MD 20850,			USA

FEATURES
 ... source
 Medical Center Dr., Rockville, MD 20850, USA
 Location/Qualifiers
 1..10908
 /organism="Chlamydia muridarum"
 /db_xref="taxon:83560"
 /note="synonym: Chlamydia trachomatis MoPn"
 241..495
 /gene="TC0265"
 241..495
 /gene="TC0265"
 /note="hypothetical protein; identified by Glimmer2;
 putative"

```

putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAE39134.1"
/db_xref="GI:7190306"
/translation="MIIVGNLRDPSKLYFFLDPTTHVFEKSLSKTRNNKKGYSLFQI
FWGQKSGCFHFHKYKVSMTFVPYKVVGVFVQRLDHLKK"
gene
426..1016
/gene="TC0266"
CDS
426..1016
/gene="TC0266"
/note="hypothetical protein; identified by Glimmer2;"

```

```
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39135.1"
/db_xref="GI:7190307"
/translation="MVCPLOGCGLLRSEASSSEEVNODPLLDALISYVRVCLAVAK
HTESAKYIHNQRKCHRGVCKGKVRFLCCCAEKEWKEGKITGLWQEHRLHG
LITGYAVHLGVCWENMLLAGETLSPGDQOLNALENSITWMLMLDSSDGKVDG
LFSAMQDLFAQYKEVKGIGNLADAVLRHFCNRT"
complement(1104..3707)
/gene="TC0267"
complement(1104..3707)
/gene="TC0267"
/translation="MRPDHVNLCCLCATILSPTAILFGQDLKSLITKNPNSIVCT
FLEDCMTENFSPALLSHARQDDPLYIIGNTHWFSVNLHSTNEERFLKERGDLISIQD
PREFLSTCSSTSPSILYHKGQSLFRVNNHNSFYRNHSEGGGALSTDALFLQFH
NYLFTFNENSSAKNGGALQATLSLSRVNLSLSPSRNRLNGAICCONLCGNGV
NPLFTNNSALNGAICCONLEONLSEKCLSLAYNOETFLSGNKAKEGGALYKHMV
LRHNGPVFVNNASKGLGAIAIQSGSLSIAGGGSVLFQNNCHFSQGVNRNAIL
KNNALLSLEARHGDLIFEDPTIQVSVSPFSTTSALTPLRIQTNRNRAVIFSEENL
KEEKEANLISKIQOPIELQSGCLVKDRVILSAPLSQAQALIVMDVGTSLTSSD
LKLTLTSLHSIDTENSYSIQSPTLSIKIPLNSHENFVENVLLSKQKIDPIL
SLPGLPDLPLDGLNLSHFGYQGMNWSQTSORETLVANWTANSYIPIPERQSAL
VANTLWNTYSDQAVQSMINTAQCGALFTGWGSAVNLSFYSHGNSKSDNDKNHRS
LGYLFIETHSDHSLFCLAGLFGSSDSFVTSADTTSYIAAIOATSLIKISA
QACYNESHELKTYRKSFGFGAWHVAVSGETGASIPVNSGSLFSSFSIFSKL
QCGSKQDGFESRGEARAFADSFNTISLSPVIAFEKSKQTRNYHFLGAYIQDLK
RVESGPVTLKNSVTWDAPMANLDSRAWMFLTNQRLHRFQTLVNNYSMLRGQSY
YSIDLGTNYR"
3956..5800
/gene="TC0268"
3956..5800
/gene="TC0268"
/translation="MFYFLGFWMGKGVGGSHSDYPIPSHNGDGESEKNSDSTSS
KVNATYSLQAPSTNDNSVPSYVDVDTLIESGSSRHVVIKKSITETEAHRES
SVEGHSSRGIFGRLOGLRLARVGEAVNTVGSIFPQAGAEQRTGKARTKISY
SASRGLRMFTDWRVYRVLHRNPPMDGLFAKLDADAEDMAATKEYYSNLFKGAAD
RETIEHCQVAKNWRKARDLRDMGAARKFLRDPFGKDPKYKGTLPGEYTVGNTMFY
DPGVSKLSEVTQFWLDMKLSDAVLSANIOGLRARFVLNOSTIPOLESLEPERFKL
ESACDARASLEAGWIKGEPKPKAQRFRFVRESNLELSFGSGESAPLSARV
SGLAAGAIRACRDCRKGKSLKDKLSSELNLABELIRFTDMGIERDDPDGNYNI
PWENWRTGPVIEGGEAHIYETWMPQESFEQYEVMDGLEERDRFASQQHYQV
PRRSLNYTPRFREYDPRNARSYDVPYPPQNEVEEMHVTGMRSSVYACFVAG
MKNYIVSQPEQIIPNSEQVEQLFQELINDGQIIQELAKIWNELDNQ"
complement(5923..6192)
/gene="TC0269"
complement(5923..6192)
/gene="TC0269"
/translation="MCKYCRGLSSLIIVVLGAINIGILGATGYRNVINARILGEGTML
TOAAYVAGTAGVIAFLNFKCFKCRKRDCCSKGCGHHHTDRE"
6393..6695
/gene="TC0270"
6393..6695
/gene="TC0270"
/translation="MYRKSALERDVAVNEISVTAITEYFHYHIESYDFQIGSFLSL
CKERALLURRIDDKLAKGDPGILAGIPIGVKDNHITGVKTKTCAKMLNENVAAPD
ATVVRRLEMEGDILLKLNDEFMAGSTTRYSAFQHTNNPMDLERVPGSSGSAAPV
SARPCFALGSDTGSIRPAACFVGVGFPSYCAVSRGLYVAFGSLDQIGPLTVV
EDVALAMDAGRDIKDATTDFRGTFSOALSLEVPLKLGVPFGDLQEDCKENF
FEALVAMEROGSRITDIDLSVLKHAVPVYIVASAEATNLARFDGVRGHRCAQADN
MOENYARSRGEGKQWTRILLIGNVLSAERONIFVYKCTAVRATLIEAFQAFEC
DVATVPCASPAIRDDVDLPVSLYLODITTVAVNLAYLPAISVPSGLSKEGLPLGVQ
FIGRKGADQIQCVGSFQSHSQIKQLYPKAVNGLFDGME"
8184..9650
/gene="TC0272"
8184..9650
/gene="TC0272"
/translation="MGIATHEWSEVIGLEHVHVELNTASKLFSFARNHFGDEPNTNISP
VCTMGALPVINKDAVRKAVFCGAVEGDVALFREDRKSIFYPDSPRFQITQYEH
PIVRGCVRAIVEEGEKTFLAOTHELDGMLKHFGDAGVDYNRAGVPLRIETVSKP
CMFSAEDAVANALVSTLSYIGISDCNMEGVSVRVDVNSVRPKSKELRNKVEIKN
MNSTFMAQALEAEKHQIEBYLSHPNEDPKVVPATYRWDPEKKKTVLMRLKERAE
DYMTFVDPDLPLQITESTYIDEVQRTLPDELPHSKMYIITDFDLAEDLAMILYSDRHT
AHFFETATMSCKNTRALSNWITVEFAGRCATKTLPTGILPEWVAQLVNFIDRGVI
TGKIAKTEADKMSVSGESPELIRHRPSSLPTDDTHALRAIVKEVVAQNAASVEDYK
NGKALGLFLVGQIMKRTGKAPKPRVNEILLAAARDA"
complement(9731..10834)
/gene="TC0273"
complement(9731..10834)
/gene="TC0273"
/translation="MVKIMAPITPTTSQVKGLLSRFLTAPDRHPKRLRYVDISLIAI
SILCIVSTIILWTOGSLAFAIPALAGLVTLLVSDLAESPKSEVADTVAASVL
PILITGTAGLMSAIAVGGAVILANFLMGSMTLGCFALMSLHKVTVYVLSNRSOW
QKONKIOESAAWENKLPKESKESQTSQTSVRYSSLARKDKTRNKGCMNGKQVPA
SIANTERSEEVHLSQRLQKELFPNTNKKELPNTKSLIHLTLNRRSPSGSOS
DDVYTPRAGLSAETSALDISGTSSTSKTSTPKAKRRVRRSSRERNARHNRN
KEDHRQNEESSDDSDSPLSPRRKKYRSRPK"
3210 a 1860 c 2587 g 3251 t
```

ORIGIN

Query Match	48.9%	Score 23:	DB 1:	Length 10908;
Best Local Similarity	74.4%	Pred. No.	2.9e+02;	
Matches	29;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	8	atgcggcgcatattcggccctgatcggccgcaagcttgg	46	
DB	4562	AAGCTGGCATATCTTCAGCCCTCATCGGCATCAAGCTTTG	4524	

RESULT	9
LOCUS	BACCOTJABC
DEFINITION	BACCOTJABC 2013 bp DNA BCT 15-JUN-1996 Bacillus subtilis cotJABC operon, polypeptide(s) affecting spore coat composition, cds.
ACCESSION	L38014
VERSION	L38014.1
KEYWORDS	GI:1377748
SOURCE	Bacillus subtilis (strain 168) DNA.
ORGANISM	Bacillus subtilis
	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus. 1 (bases 1 to 2013)
REFERENCE	Henriques, A.O., Beall, B.W., Roland, K. and Moran, C.P. Jr.
AUTHORS	Characterization of cotJ, a sigma E-controlled operon affecting the polypeptide composition of the coat of Bacillus subtilis spores
TITLE	J. Bacteriol. 177 (12), 3394-3406 (1995)
JOURNAL	

FEATURES	source	Location/Qualifiers
33260532		
		1. .2013
		/organism="Bacillus subtilis"
		/strain="168"
		/db_xref="taxon:1423"
gene		326..574
		/gene="cotJA"
CDS		326..574
		/gene="cotJA"
		/note="putative"
		/codon_start=1
		/transl_table=11
		/protein_id="AAB17588.1"
		/db_xref="GI:1377749"
		/translation="MKDQPFPTVKSYTPFHSREDPCPPIGKYKYYRTPNLYMTFOE HMQQSPMLAKRGTLWKDLDFENPYRGDAHGKKG"
gene		519..821
		/gene="cotJB"
CDS		519..821

BASE COUNT
ORIGIN

Query Match	48.1%	Score 22.6	DB 2	Length 1013
Best Local Similarity	68.9%	Pred. No. 5e+02		
Matches 31	Conservative	0	Mismatches 14	Indels 0
Gaps	0			

RESULT	10
AC017738	
LOCUS	AC017738
DEFINITION	DNA
	17103 bp HTG
	10-DEC-1999
	*** SEQUENCING IN PROGRESS ***
	, in ordered pieces.
AC017738	
VERSION	AC017738.1
KEYWORDS	GI:6554261
SOURCE	HTG; HTGS_PHASE2.
ORGANISM	fruit fly.
	Drosophila melanogaster
Eukaryota;	Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota;	Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha;	Ephydroidea; Drosophilidae; Drosophila.
	1 (bases 1 to 17103)
ADAMS,M. and Venter,J.C.	
Direct Submission	
Submitted (10-DEC-1999)	Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA	
This sequence was identified as CDW:10211936 by the submitter.	
For more information on this record e-mail to fly@celera.com.	
* NOTE:	This is a 'working draft' sequence.
*	This sequence will be replaced
*	by the finished sequence as soon as it is available and
*	the accession number will be preserved.

```

FEATURES
  source
    Location/Qualifiers
      1. .17103
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
BASE COUNT      5111 a 3705 c 3714 g 4573 t
ORIGIN
Query Match      48.1%; Score 22.6; DB 59; Length 17103;
Best Local Similarity 68.9%; Pred. No. 3.8e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1 agctataatgcggccgcatattcgccgtgatogggccgaagcttg 45
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12160 AGCGATCAGCGCTGCTCCCTGTTCCGCCAGCTCGCTGGAAGTTG 12204

```

RESULT	11
LOCUS	AF012532/c
DEFINITION	12-AUG-1997 BCT 18622 bp DNA Bacillus subtilis strain 168 trpC2 YefA (yefA) gene, partial cds, and YefB (yefB), YefC (yefC), YeeA (yeeA), YeeB (yeeB), YeeC (yeeC), YeeD (yeeD), YeeE (yeeE) and YeeF (yeeF) genes, complete cds.
ACCESSION	AF012532
VERSION	AF012532.1
KEYWORDS	GI:2318058
SOURCE	Bacillus subtilis.
ORGANISM	Bacillus subtilis. Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/staphylococcus group; Bacillus.
REFERENCE	1 (bases 15361 to 18622) Henriques,A.O., Beall,B.W., Roland,K. and Moran,C.P. Jr.
AUTHORS	Characterization of cotJ, a sigma E-controlled operon affecting the polypeptide composition of the coat of Bacillus subtilis spores J. Bacteriol. 177 (12), 3394-3406 (1995)
JOURNAL	95286532
MEDLINE	2 (bases 1 to 18622)
REFERENCE	Borriess,R. and Schroeter,R.
AUTHORS	

RESULT	11
LOCUS	AF012532/c
DEFINITION	12-AUG-1997 BCT 18622 bp DNA Bacillus subtilis strain 168 trpC2 YefA (yefA) gene, partial cds, and YefB (yefB), YefC (yefC), YeeA (yeeA), YeeB (yeeB), YeeC (yeeC), YeeD (yeeD), YeeE (yeeE) and YeeF (yeeF) genes, complete cds.
ACCESSION	AF012532
VERSION	AF012532.1
KEYWORDS	GI:2318058
SOURCE	Bacillus subtilis.
ORGANISM	Bacillus subtilis. Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/staphylococcus group; Bacillus.
REFERENCE	1 (bases 15361 to 18622) Henriques,A.O., Beall,B.W., Roland,K. and Moran,C.P. Jr.
AUTHORS	Characterization of cotJ, a sigma E-controlled operon affecting the polypeptide composition of the coat of Bacillus subtilis spores J. Bacteriol. 177 (12), 3394-3406 (1995)
JOURNAL	95286532
MEDLINE	2 (bases 1 to 18622)
REFERENCE	Borriess,R. and Schroeter,R.
AUTHORS	

TITLE The 55-58 degree segment of the *Bacillus subtilis* chromosome, a region spanning from the *purA* gene cluster to the *cotJ* operon

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 18622)

AUTHORS Borriass, R.

TITLE Direct Submission

JOURNAL Submitted (03-JUL-1997) Borriass R., Humboldt University, Institute of Biology, Chausseestraße 117, Berlin, Germany, D-10115

FEATURES Location/Qualifiers

1..18622

`/organism="Bacillus subtilis"`

`/strain="168 trpC2"`

`/db_xref="taxon:1423"`

`/map_55 to 58 degree`

`<1..81`

`/gene="yefA"`

`<1..81`

`/gene="yefA"`

`/note="similar to E. coli hypothetical RNA methyltransferase; SwissProt Accession Number P55135; similar to Chlamydia trachomatis Hom1; SwissProt Accession Number P55137"`

`/citation=[2]`

`/codon_start=1`

`/transl_table=11`

`/product="YefA"`

`/protein_id="AAB66480.1"`

`/db_xref="GI:2318067"`

`/translation="AKRLRRSPADNCGDEAAGSVCL"`

`complement(227..1140)`

`/gene="yefB"`

`complement(227..1129)`

`/gene="yefB"`

`/citation=[2]`

`/codon_start=1`

`/transl_table=11`

`/product="YefB"`

`/protein_id="AAB66472.1"`

`/db_xref="GI:2318059"`

`/translation="MYNEGVSPSLARKNATDKWGHSSIRKILENPHYTGTLQOOR EYPSVTSRRRSNLENQIIIEKSHEPIPLEDFLIYQEIILITRRRPPQAEVHLFT NTAVCDCRRSHFKNRKGYVCGSYNKHGLKACSDHYVEIDLTKVITNLKYLK FSKNYEKLSEKALKYKEKIESKIEINEKLNDRKROKSLVSIANGVTSKEDYQL AINTNEDISNETTLHQLSKDLEQKIEKELIDFKKSLDKFMEKSGTLTPEMLHLVD EIEVHANGTIEINRPREPTVPSA"`

`1024..1445`

`/gene="yefC"`

`1024..1030`

`/gene="yefC"`

`1068..1445`

`/gene="yefC"`

`/note="similar to Enterococcus faecalis plasmid PY117 orf4 encoded by GenBank Accession Number D78257"`

`/citation=[2]`

`/codon_start=1`

`/transl_table=11`

`/product="YefC"`

`/protein_id="AAB66473.1"`

`/db_xref="GI:2318060"`

`/translation="MSVAFPLRASCEGLVTPSLYNPLAIAKPPFAIYSEKILIFTIF SAAFEERDMIVERTQEGKLAQONPDFREGPKFTKQIINHALLTENHSYKQVED MTGISVTLVRAKKKAAEAING"`

`complement(1133..1140)`

`/gene="yefB"`

`1505..4159`

`/gene="yefA"`

`1505..1512`

`/gene="yefA"`

`1520..4159`

`/gene="yefA"`

`/note="similar to Methanococcus janaschii plasmid pURB801 hypothetical protein MJEC502 encoded by GenBank Accession Number L77119"`

`Number L77119"`

`/citation=[2]`

`/codon_start=1`

`/transl_table=11`

`/product="YefA"`

`/protein_id="AAB66474.1"`

`/db_xref="GI:2318061"`

`/translation="MALIDLEDKTAETVNRHDSDFLYELLGVVDVPRATITRLKKN ONLTKRVGEVHLKKNVWFKAQKGLKDALIDIEQVEYLSAKPRYLLVTDYDGLAK DTKLEALDVFPELPQYFDFFLAWGIEKVEFEPENPADIKAEARFIYDVRKEN NIETNRGLDLFLIRLFCFAEDTDFKRNFTNLKTLTEEDGSNLKLFADLFI LDNRDDVPYSYKEFPYVNGQLFTEPTELEFSAKRKLIEGELLNWKINPDIF GSMIOAVASEBSRSYLGMYHTSVPNIMKIPFLDKLQNSFLDAYDDYTKLENLLTR IGKTFDPACGSNELLITYKELRRMEINLIKLOELLGEYLYVPSVTLSQFVGTIEI EDPAHDVAKLSLWIAEHQMEELKNEVHNVRPTPLHTAGDIRCANAIJVENTVECP AQSEEVYFNGPPTLGSKKKCKHSDMSLIFGKVGKGMLODITISAWFTGAKYAST TNARVAFVSTNSVTQGVQSILMNELFPGIQINFANKSEKANNKNAFVAIVIVG FGPLDTKVNRYLFVDETKLVSNISPYLTGDELIVSSRCKPISDLPKLHFGMNPDG GGLLFTTEVDAINKYPELVYFKEFGISGVFINGGLRCKLWNEAKYKIKSNPLI QERTISKNHREKSTDKGNKLTLPKFRDTHETTNYSIVPVSSENRFYIPMGLI GADPILSNLIYVDAEYLLGLMSBHMHTWKAVAGRLKTDYRYVYSGLCYNTFPPI ELSTRKNEIEEALIEILDREEGGTIAELYNPSTMPIELKVAHEKLDGIVERAYRQ KQFESDEERLEVLKLYOEMTER"`

`4152..4161`

`/gene="yefB"`

`4152..5967`

`/gene="yefB"`

`4171..5967`

`/gene="yefB"`

`/citation=[2]`

`/codon_start=1`

`/transl_table=11`

`/product="YefB"`

`/protein_id="AAB66475.1"`

`/db_xref="GI:2318062"`

`/translation="MENIEINYNOTGSKTKTNEYGMREMOARAFKRNQYLLVRAP PASGRALMEIGLIDKLINOGLRVIVAPERSIGSSFKNTDLKSYGFFENKVKYDPNR NLTVGDNKSVKSVRFMESDDQVLICTHSTLRFAEKIDDKAFNCLLAIDEFHVYS ADVNSRLGELLRSIHNSSAHIVAMTSGSYFGSDVPILLPEDELFOKVTYVYEQLD GYEYKRGIGYGHIIYVQRHSDGKTLADIADVDQGVGNVAVLRNEALDDLDIIL MIGVEYODDDTGIIYVQRHSDGKTLADIADVDQGVGNVAVLRNEALDDLDIIL ALGMAKGFDFWPCFCEHTLVGYRGSLEIVQIIGCTRDSYNTKYAQTNLAMPDAK DEVVTYVNTWKAIASALLMEQVLTDFDFKRRNESEKSTTGELFVGLKEPSTE NRKLIENDINDLAKTCKIHKSFTGEVDPKVLNKLVPKIQKVPVPLNSDEIEE VRQVILDTVMKGAKEVVGSKFIRWADKFINIEDLDINLDSINFPQAFEVLSKE LNSPVKLQIETIDAKRINFEDELVFWPKV"`

`6064..6069`

`/gene="yefC"`

`6064..7228`

`/gene="yefC"`

`6080..7228`

`/gene="yefC"`

`/citation=[2]`

`/codon_start=1`

`/transl_table=11`

`/product="YefC"`

`/protein_id="AAB66476.1"`

`/db_xref="GI:2318063"`

`/translation="MASNRHSINEIMSOLFQITPPPKTKQAQDPEVEKELEII EFWENGREPQVPTDLTERSLSRLIGIRKDPORMEYLYKEYDEIGLLEVRKSLTIP KISSIDDLKLSGLSALLGDNLNNDITRSIFDTSLQKVTYMPYVAKRKKTKDEKFE ELFKKCHKEITEGKRKLITLTKNEQDIQNSFYILKGVLLVYGVGKAKGKTNRAL RCIPIENGTESDMLLSLSAELYKHGRVTDNEDTLIDNVREDDVSTGFYVYLSLSTD PQISSIKNLYKIGFTTGSVENRIRNAENQSTLYLAPVEITTYQVFNNAKSKFETAIH HALENNLDVSLGANGKAPVFWKFWVTLDDLOAVIDEIVMMVHLVD"`

`complement(7490..8180)`

`/gene="yefB"`

`complement(7490..7795)`

`/gene="yefB"`

`/citation=[2]`

`/codon_start=1`

`/transl_table=11`

`/product="YefB"`

`/protein_id="AAB66477.1"`

/db_xref="GI:2318064"
/translation="MTVRKELIKQINLTITIVIKTINOKNPTMVKNIILKRYEAEKFI
LQSSDKFEDLSKVNKLDITLTRYAVLESANDYMNMLKEMHTEKLIKKEYDETTQS"
complement(8175..8180)
/gene="yeeE"
8277..8298
terminator
/note="rho independent terminator; deltaG= -10.2kcal/mol"
/citation=[2]
complement(8308..9470)
/gene="yeeE"
complement(8308..9417)
/gene="yeeE"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yeeE"
/protein_id="AAB66478.1"
/db_xref="GI:2318065"

/translation="MPAGVPYNAVNGVTLKEGLQKFAKVLPKPYGTSSGRRTPAPH
VPVTVYKGEHFARWSRKKVLPNIITKGYTITDNYGRITSVKADLOLGEAKRN
QVATWAGKPDQRKDDGGHLIATQFGSGGFDNIVPMNSQINRSGGKWKVEMGEWA
KALSKPKPKVAQVTEPVYSGDSLPSYFDVTYKIGSRKEISVSIKISLIGVEMETRK
MODLYLIGELKNDIIPGEWTIKIYLAEVLDDSTAVLHFRTPENNNQIIYSQDIPSHY
NYSKDLFKLLRELFEELREKRLGILPELDEDDKKEFVSYLGL"
complement(9464..9470)
/gene="yeeE"
complement(9541..10791)
/gene="yeeE"
complement(9541..10791)
/gene="yeeE"
/note="similar to Bacillus subtilis orf94 hypothetical
59.7 kDa protein in CWLA-CISA intergenic region: SwissProt
Accession Number P45942"

Query Match 48.1%; Score 22.6; DB 1; Length 18622;

Best Local Similarity 68.9%; Pred. No. 3.7e+02;

Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 3 ctataatgcgcgcattatgcacatgacgagcgcgcaagcttga 47

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 17005 CCAAAATGAGTCCAAATATGACGATGAGCAGCGCAACTCTGA 16961

RESULT 12

AC009915

LOCUS

AC009915 101962 bp DNA HTG 10-FEB-2000
Drosophila melanogaster chromosome 3 clone BACR09L14 (D751) RPOCI-98
09.L.14 map 94A-94A strain y; cn bw sp, *** SEQUENCING IN PROGRESS
***, 84 unordered pieces.

ACCESSION

AC009915.4 GI:6957926

VERSION

HTG; HTGS_PHASE1.

KEYWORDS

fruit fly.

SOURCE

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 101962)

Celniker, S.E., Aghayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,

Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,

Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

Hinkley, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,

Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R.,

Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B.,

Richards, S., Sethi, H., Swirskas, R.R., Wan, K.H., Webster, D.,

Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.

Sequencing of Drosophila melanogaster

Unpublished

2 (bases 1 to 101962)

Celniker, S.E., Aghayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,

Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,

Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

TITLE

JOURNAL

COMMENT

Direct Submission

Submitted (08-SEP-1999)

Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Feb 11, 2000 this sequence version replaced gi:6598737.

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive web site (<http://www.fruitfly.org/sequence/>) or send emailto bdg@fruitfly.berkeley.edu. All contigs in this submission meet

the following cutoffs: length >= 200 bases.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 84 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1

1017: contig of 1017 bp in length

1018 1097: gap of unknown length

1098 1838: contig of 741 bp in length

1839 1918: gap of unknown length

1919 2723: contig of 805 bp in length

2724 2803: gap of unknown length

2804 3804: contig of 1001 bp in length

3805 3884: gap of unknown length

3885 4499: contig of 615 bp in length

4500 4579: gap of unknown length

4580 5031: contig of 452 bp in length

5032 5111: gap of unknown length

5112 6264: contig of 1153 bp in length

6265 6344: gap of unknown length

6345 6971: contig of 627 bp in length

6972 7051: gap of unknown length

7052 7722: contig of 671 bp in length

7723 8423: gap of unknown length

8424 8503: contig of 621 bp in length

8504 8988: contig of 485 bp in length

8989 9069: gap of unknown length

9070 9640: gap of unknown length

9641 9720: contig of 677 bp in length

9721 10397: contig of 600 bp in length

10398 10477: contig of 600 bp in length

10478 11077: gap of unknown length

11078 11156: gap of unknown length

11157 11722: contig of 566 bp in length

11723 13112: contig of 1310 bp in length

13113 13192: gap of unknown length

13193 14134: contig of 942 bp in length

14135 14214: gap of unknown length

14215 14808: contig of 594 bp in length

14809 14888: gap of unknown length

14889 15796: contig of 908 bp in length

15797 15876: gap of unknown length

15877 16944: contig of 1068 bp in length

16945 17024: gap of unknown length

17025 18221: contig of 1197 bp in length

18222 18301: gap of unknown length

18302 19185: contig of 884 bp in length

19186 19265: gap of unknown length

19266 19953: contig of 688 bp in length

19954 20034: gap of unknown length

20035 21121: contig of 1087 bp in length

21122 22100: gap of unknown length

22101 22200: contig of 920 bp in length

22201 22200: gap of unknown length

Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H., Zumbstein, E., Yoshikawa, H. and Danchin, A.

The complete genome sequence of the gram-positive bacterium *Bacillus subtilis* Nature 390 (6657), 249-256 (1997)

98044033

2 (bases 1 to 213190)

Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.

Direct Submission

Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

FEATURES

Location/Qualifiers

1..213190
 /organism="Bacillus subtilis"
 /strain="168"
 /db_xref="taxon:1423"
 33..740
 /gene="ydfs"
 33..740
 /gene="ydfs"
 /function="unknown"
 /note="similar to hypothetical proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12373.1"
 /db_xref="GI:2632867"
 /db_xref="SPTREMBL:P96697"
 /translation="MIELEVIRTVASFGLLIAERILGKQTISQMTIDFPIAATLG
 AIAAGAYNTSKPHMAISFIVLTIFLISFKIKRKRKFAGDPTVLQNGKI
 LESNRKMYTLDYLNQLRKEIEFIEVLAILETNGQLTLRKQFRTVKODLM
 IAVNQERPLIELMDGETIENLNKQRLTESWLEELRKRDIKVKETVYAVLLGNGD
 IYDQYKDHISVPMDEK"
 complement(755..1186)
 /gene="ydfv"
 complement(755..1186)
 /gene="ydfv"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12374.1"
 /db_xref="GI:2632868"
 /db_xref="SPTREMBL:P96698"
 /translation="MDFEKIRKWLIEITNEYKQSDFTWLVKYPKAPHEFFDSEATFVY
 DFQDEYNFIVEMPGVVEELTIRLSKTLQIKGTITPVFPAEMEVLRERYGEI
 BRIOLPEAEPTHLQIOLNGLLHISYPRQVEIVAFNKL"
 complement(1199..1441)
 /gene="ydgA"
 complement(1199..1441)
 /gene="ydgA"
 /function="unknown"
 /note="similar to hypothetical proteins from B. subtilis"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12375.1"
 /db_xref="GI:2632869"
 /db_xref="SPTREMBL:P96699"
 /translation="MPYOINIANIKINGVTONGNIDVCGPTVHNSHTANSKYEGANFSL
 GDLSPTSILNTGNDSDVSDQDQIGNFAPISNQI"
 complement(1455..1727)
 /gene="ydgB"
 complement(1455..1727)
 /gene="ydgB"
 /function="unknown"
 /note="similar to hypothetical proteins from B. subtilis"
 /codon_start=1

gene

CDS

gene

CDS

gene

CDS

gene

CDS

/transl_table=11
 /protein_id="CAB12376.1"
 /db_xref="GI:2632870"
 /db_xref="SPTREMBL:P96700"
 /translation="MPSTVINLYLKINSISGNSITIGEAAYNSPTNNKSOQNTSS
 FGDSPTSESVNENFLNDPDVNDQTSIGNSDTSNNAPIAPPILD"
 2026..2613
 /gene="ydgC"
 2026..2613
 /gene="ydgC"
 /function="unknown"
 /note="similar to hypothetical proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12377.1"
 /db_xref="GI:2632871"
 /db_xref="SPTREMBL:P96701"
 /translation="MKSQGVESKRLLKAAANEFVSRGFHDKAYSEIVKAGETQPSF
 YLYFQSKAEAFIAELITDFHSRVKRLTESLLENGLNTEDYSKRVLLAVETVFGFLDED
 KDLTKIGFLNPEAKQMKDLAMVLKENLEAEORLGYFHSLEDMETVAECIVGMIEHL
 TESPELLGKIDPASLAAYVNLIIYGMPLKGNDR"
 join(2610..2954,3074..3091)
 /gene="ydgD"
 2610..2954
 /gene="ydgD"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12378.1"
 /db_xref="GI:2632872"
 /db_xref="SPTREMBL:P96702"
 /translation="MISIMKVSILAVFVLVAGGIKVSVPFQVEHRRHYOYPLWFLTV
 TGILEIAGALAMTAGIWNRYAAGVLFVVLVAGALHAHMERARQSVINAIQAMICL
 IVSIIIMTIGSYT"
 3074..3091
 /gene="ydgD"
 3117..3590
 /gene="ydgE"
 3117..3590
 /gene="ydgE"
 /function="unknown"
 /note="similar to hypothetical proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12379.1"
 /db_xref="GI:2632873"
 /db_xref="SPTREMBL:P96703"
 /translation="MAFTLEDMTTEEEFEAFRGMVSQVONFAKONITSGTWTEKEAFKSE
 QAYENMIPNGRDSNHYFWNITNDQGERMGLWLYADPLHPQKEAFYSGFLYEAFRG
 KGLAQALQTLDERARELGAERLALHVFHAHNTAVLYQKMGVAMTIRMKQLC"
 complement(3705..3722)
 /gene="expZ"
 complement(3705..3722)
 /gene="expZ"
 complement(3705..3722)
 /gene="expZ"
 complement(3750..5393)
 /gene="expZ"
 complement(3750..5393)
 /gene="expZ"
 /codon_start=1
 /transl_table=11
 /product="ATP-binding transport protein"
 /protein_id="CAB12380.1"
 /db_xref="GI:2632874"
 /db_xref="SWISS-PROT:P39115"
 /translation="MKEIVLTNVSVEVKDQTVFKHVNASVOOGDIIIGIKNGACKS
 TLHLHLNDLAPAQGOILRKDKIALVEQETAAYSPADOTPAEKKILKXWHVPLRDPH
 QLSGGEKARKLAKLAGSEADALLLEPTNHLDEKSLQFLIOLKHHYNGTVILVSHDR
 YFLDEKATKWSLEDTLFEKNGYSYMKRFRKKRLTOOREYKQKQWVERLEAQM
 GLASWSEKAHQAKTKEGKQYHRVAKRTDAQIKSKRLEKELEKAKAPVTPYT
 VRESIDTHTKGRFLVQNVTKAFGERTLFKNANFTIQHGEKVAIIGPNSGKTTLL
 NIILQGTAEQSVWSPSANGYLTQETPEELFENFETKARGHVNLMRH
 LGFTAAQWTEPIKHMMSGGERVKIIMAYILLEKDVLLLEPTNHLDPRLQLEETLS
 QYSGTLLAVSHDRYFLEKTNLSKLVISNNGIEKQLNDVPSERNEREELRLKLETERQE
 VLKGLSFMPNDQAFNELTKRIKELDHQDKKD"
 complement(5428..5450)

terminator

gene

CDS

terminator

gene

CDS

terminator

[illegible]

	/map=9A10-94A12"	
	/db_xref="FLYBASE:Fban007086"	
	/db_xref="FLYBASE:Fpan0038937"	
CDS	join(98734..98776,98909..99054)	.99209,.99474,99546..996
	100055..100449,100525..100801)"	
	/gene=Cg7086	
	/note="Cg7086 gene product"	
	/codon_start=1	
	/db_xref="FLYBASE:Fpan007086"	
	/db_xref="FLYBASE:Fpan0038937"	
	/protein_id="AAFS5962.1"	
	/db_xref="GI:7300820"	
	/translation="MSVSNMGMFMFHSIQALVRPSPHSSISAEDRTSMASCICKA	
Query Match	48.1%;	Score 22.6; DB 30; Length 22615;
Best Local Similarity	68.9%;	Pred. No. 2.7e+02;
Matches 31; Conservative	0;	Mismatches 14; Indels 0; Gaps 0;
Oy 1	aactataatgcggccagcatttcgcccgcatgcgcgaacgttg	45
Dbb 97887	AAGCATTCAGGCGCTTGCCCTGTTCGGCACGCCTCCGGGAAGAATTGG	97931

AVI/GLNA/C	AVI/GLNA	1950 bp	DNA	BCF	06-MAR-1995
LOCUS	A.vinelandii	glutamine synthetase (glnA) gene, complete cds.			
DEFINITION	A.vinelandii	glutamine synthetase (glnA) gene, complete cds.			
ACCESSION	M57275				
VERSION	M57275.1	GI:142308			
KEYWORDS	glutamine synthetase.				
SOURCE	Azotobacter vinelandii DNA.				
ORGANISM	Azotobacter vinelandii				
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Azotobacter.				
AUTHORS	1 (bases 1 to 1950) Toukdarian,A., Saunders,G., Selman-Sosa,G., Santero,E., Woodley,P. and Kennedy,C.				
TITLE	Molecular analysis of the Azotobacter vinelandii glnA gene encoding glutamine synthetase				
JOURNAL	J. Bacteriol.	172 (11),	6529-6539	(1990)	
MEDLINE	91035268				
FEATURES					
source	Location/Qualifiers				
	1..1950				
	/organism="Azotobacter vinelandii"				
	/isolate="UW136"				
	/strain="Op"				
	/db_xref="taxon:354"				
	/clone="pLW50"				
	446..>1950				
	/gene="glnA"				
	/evidence=experimental				
	446..1950				
	/gene="glnA"				
	478..1881				
	/gene="glnA"				
	/EC_number="6.3.1.2"				
	/codon_start=1				
	/transl_table=1				
	/product="glutamine synthetase"				
	/protein_id="AA62673.1"				
	/db_xref="GI:142309"				
	/translation="MSKSLDIKHDYKWLDRPTDJKKQOHVTMPARDVDFFRY GKMDSGIAGWKGIENASDMLIKHDYKWLDRPTDJKKQOHVTMPARDVDFFRY PRATARRAEELKSTGIGDTAFEGPEPEFIFEDVKYKSDISGMRIFESQAMWTD ADPEGNTKGRHHPGVGGYFPVPVDHDEIRITACNALIEEMGLKVEVHNHEVITAGN EIGFSENTLVAKADVEQTLKVCVHNADAYGKVTPEMPKPLDNGSGMHVHSIKAD GKNSFAGEGYAGSDTALYFIGGIRKGAALNGFTNSTSKYKRLVPGFAPVYMLAS ARNSASATIRIPYVNSPKARIRIEMARFPDPSNPLATAPALIMAGDIOGKTIHSGDAAD KNLYDPEPEAKETPOYCGSLKEALEELDGRAFLLTKGVFSDDFTDAYIELKSEETI KVRTEVPLEDLYSY"				

Mon Mar 5 13:21:08 2001

us-09-101-423b-8.rge

Page 15

Query Match	47.2%	Score 22.2:	DB 2,	length 1950:
Best Local Similarity	77.1%	Pred. No	7e+02;	
Matches	27;	Conservative	0;	Mismatches 8; Indels 0; Gaps 0;
QY	2	gctataatgcgagcgcatatcgcgcctatcgcc	36	
db	200	gctatTAATcccccctCTCTTTGcctACcc6cc	166	

Search completed: March 4, 2001, 12:04:02
job time: 5617 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 09:48:30 ; Search time 1300.69 Seconds
(without alignments)
242.438 Million cell updates/sec

Title: US-09-101-423B-7
Perfect score: 45

Sequence: 1 aatccaagcttgccgcgat.....cgaatatgcgcgcattat 45

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 segs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb.est1:*
2: gb.est2:*
3: gb.est3:*
4: gb.est4:*
5: gb.est5:*
6: gb.est6:*
7: gb.est7:*
8: gb.est8:*
9: gb.est9:*
10: gb.est10:*
11: gb.est11:*
12: gb.est12:*
13: gb.est13:*
14: gb.est14:*
15: gb.est15:*
16: gb.est16:*
17: gb.est17:*
18: gb.est18:*
19: gb.est19:*
20: gb.est20:*
21: gb.est21:*
22: gb.est22:*
23: gb.est23:*
24: gb.est24:*
25: gb.est25:*
26: gb.est26:*
27: gb.est27:*
28: gb.est28:*
29: gb.est29:*
30: gb.est30:*
31: gb.est31:*
32: gb.est32:*
33: gb.est33:*
34: gb.est34:*
35: gb.est35:*
36: gb.est36:*
37: gb.est37:*
38: gb.est38:*
39: gb.est39:*
40: gb.est40:*
41: em.estda:*
42: em.estfun:*
43: em.esthum1:*

44: em.esthum2:*
45: em.esthum3:*
46: em.esthum4:*
47: em.esthum5:*
48: em.esthum6:*
49: em.esthum7:*
50: em.esthum8:*
51: em.esthum9:*
52: em.esthum10:*
53: em.esthum11:*
54: em.esthum12:*
55: em.esthum13:*
56: em.esthum14:*
57: em.esthum15:*
58: em.esthum16:*
59: em.esthum17:*
60: em.esthum18:*
61: em.esthum19:*
62: em.esthum20:*
63: em.estin1:*
64: em.estin2:*
65: em.estin3:*
66: em.estin4:*
67: em.estov1:*
68: em.estov2:*
69: em.estp11:*
70: em.estp12:*
71: em.estp13:*
72: em.estp14:*
73: em.estp15:*
74: em.estrol1:*
75: em.estrol2:*
76: em.estrol3:*
77: em.estrol4:*
78: em.estrol5:*
79: em.estrol6:*
80: em.estrol7:*
81: em.estrol8:*
82: em.estrol9:*
83: em.estrol10:*
84: em.estrol11:*
85: em.estrol12:*
86: em.estrol13:*
87: gb.est41:*
88: gb.est42:*
89: gb.est43:*
90: gb.est44:*
91: gb.est45:*
92: gb.est46:*
93: gb.est47:*
94: gb.est48:*
95: gb.est49:*
96: gb.est50:*
97: gb.est51:*
98: gb.est52:*
99: gb.est53:*
100: gb.est54:*
101: gb.est55:*
102: gb.est56:*
103: gb.est57:*
104: gb.est58:*
105: gb.est59:*
106: gb.est60:*
107: gb.est61:*
108: gb.est62:*
109: gb.est63:*
110: gb.est64:*
111: gb.est65:*
112: em.esthum21:*
113: em.esthum22:*
114: em.esthum23:*
115: em.esthum24:*
116: em.esthum25:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estp19:*
 121: em_estp20:*
 122: em_estp21:*
 123: em_estp22:*
 124: em_estp23:*
 125: em_estp24:*
 126: em_estp25:*
 127: em_estp26:*
 128: em_estp27:*
 129: em_estp28:*
 130: em_estp29:*
 131: em_estp30:*
 132: em_estp31:*
 133: em_estp32:*
 134: em_estp33:*
 135: em_estp34:*
 136: em_estp35:*
 137: em_estp36:*
 138: em_estp37:*
 139: em_estp38:*
 140: em_estp39:*
 141: em_estp40:*
 142: em_estp41:*
 143: em_estp42:*
 144: em_estp43:*
 145: em_estp44:*
 146: em_estp45:*
 147: em_estp46:*
 148: em_estp47:*
 149: em_estp48:*
 150: em_estp49:*
 151: em_estp50:*
 152: em_estp51:*
 153: em_estp52:*
 154: em_estp53:*
 155: em_estp54:*
 156: em_estp55:*
 157: em_estp56:*
 158: em_estp57:*
 159: em_estp58:*
 160: em_estp59:*
 161: em_estp60:*
 162: em_estp61:*
 163: em_estp62:*
 164: em_estp63:*
 165: em_estp64:*
 166: em_estp65:*
 167: em_estp66:*
 168: em_estp67:*
 169: em_estp68:*
 170: em_estp69:*
 171: em_estp70:*
 172: em_estp71:*
 173: em_estp72:*
 174: em_estp73:*
 175: em_estp74:*
 176: em_estp75:*
 177: em_estp76:*
 178: em_estp77:*
 179: em_estp78:*
 180: em_estp79:*
 181: em_estp80:*
 182: em_estp81:*
 183: em_estp82:*
 184: em_estp83:*
 185: em_estp84:*
 186: em_estp85:*
 187: em_estp86:*
 188: em_estp87:*
 189: em_estp88:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	23.4	52.0	313	18	AI283439	AI283439 gb93e10.x
C 2	23.4	52.0	520	4	AA236621	AA236621 z843g09.r
C 3	23.2	51.6	625	21	AI514427	AI514427 LD40932.5
C 4	23.2	51.6	1054	134	BE036773	BE036773 MP05D02.M
C 5	22.4	49.8	705	97	AW977483	AW977483 EST189592
C 6	22.2	49.3	528	134	BE062593	BE062593 QV1-EP026
C 7	22.2	49.3	558	163	AO841408	AO841408 T136718b
C 8	22.2	48.9	266	146	W06254	W06254 T9EST278d0
C 9	22.2	48.9	267	14	AA952587	AA952587 TEN81771
C 10	22.2	48.9	313	23	AI667900	AI667900 TEN80871
C 11	22.2	48.9	373	146	W35559	W35559 T9EST278b0
C 12	21.8	48.4	395	87	AW185272	AW185272 se89c06.y
C 13	21.8	48.4	436	37	AV624476	AV624476 AV624476
C 14	21.8	48.4	610	87	AW181098	AW181098 MGA00627
C 15	21.8	48.4	615	157	AO399560	AO399560 MGA000140
C 16	21.8	48.4	635	157	AO398650	AO398650 MGA000118
C 17	21.8	48.4	640	87	AW181091	AW181091 MGA00327
C 18	21.6	48.0	541	159	AO578848	AO578848 ndxb0093F
C 19	21.6	48.0	650	7	AA438826	AA438826 LD13184.5
C 20	21.6	48.0	677	138	BE976052	BE976052 bs46f11.x
C 21	21.6	48.0	707	97	AW940801	AW940801 GH20963.3
C 22	21.4	47.6	947	193	CNS05G00	AL336561 Tetraodon
C 23	21.2	47.1	263	36	AV425024	AV425024 AV425024
C 24	21.2	47.1	266	36	AV426457	AV426457 AV426457
C 25	21.2	47.1	403	36	AV409533	AV409533 AV409533
C 26	21.2	47.1	463	36	AV422473	AV422473 AV422473
C 27	21.2	47.1	474	191	CNS03AWS	AL235765 Tetraodon
C 28	21.2	47.1	912	97	AW925378	AW925378 HVSME000
C 29	21.2	47.1	991	191	CNS02VLE	AL215915 Tetraodon
C 30	21.2	46.7	188	23	AI690000	AI690000 t32a09.x
C 31	21.2	46.7	220	91	AA491261	AA491261 UI-M-BH3-
C 32	21.2	46.7	244	137	BE943769	BE943769 UI-M-BH3-
C 33	21.2	46.7	272	91	AW491371	AW491371 UI-M-BH3-
C 34	21.2	46.7	331	144	R87445	R87445 ym89b12.s1
C 35	21.2	46.7	350	15	AI024322	AI024322 ov67g07.x
C 36	21.2	46.7	351	107	BE363902	BE363902 P11_10_10
C 37	21.2	46.7	352	24	AI768069	AI768069 w146d08.x
C 38	21.2	46.7	371	14	AB009130	AB009130 AB009130
C 39	21.2	46.7	391	36	AV420471	AV420471 AV420471
C 40	21.2	46.7	408	7	AA463976	AA463976 z86f01.s
C 41	21.2	46.7	430	105	BE222661	BE222661 hu49h10.x
C 42	21.2	46.7	448	110	BE647694	BE647694 UI-M-BH1-
C 43	21.2	46.7	463	39	AW077673	AW077673 t336p05.y
C 44	21.2	46.7	487	18	AI292116	AI292116 gm86e02.x
C 45	21.2	46.7	488	18	AI291238	AI291238 gm11b04.x

ALIGNMENTS

RESULT 1
 AI283439/c
 LOCUS 313 bp mRNA
 DEFINITION gb93e10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 IMAGE:1854570 3' similar to TR:P76904 P76904 SIMILAR TO ; mRNA
 ACCESSION AI283439
 VERSION AI283439.1 GI:3921672
 KEYWORDS EST.

SOURCE

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS

1 (bases 1 to 313)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -400P from Glibco

High quality sequence stop: 297.

FEATURES

source

1..313

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1854570"

/clone_lib="Soares_NFL_T_GBC_S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBH119W, testis NHT, and B-cellNCLCGAP-GCB1) were mixed and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driverwas PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted ofI.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo.

BASE COUNT

67 a 96 c 104 g 46 t

Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Query Match

Best Local Similarity 52.0%; Score 23.4; DB 18; Length 313;

Matches

30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY

4 ccaagcttgccgacgacgaatgagccgacatg 44

Db 75 CCCAGCTGCCGCGACGACGCGCAATGTCGCCGAGTA 35

RESULT

AA236621

520 bp mRNA

EST

07-AUG-1997

LOCUS

Z543909.1

Soares_NHMPU_S1 Homo sapiens cDNA clone IMAGE:688000 5'

Similar to TR:G1079677 G1079677 LPESP.; mRNA sequence.

ACCESSION

AA236621

GI:1860641

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS

1 (bases 1 to 520)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins

JOURNAL

Unpublished (1997)

COMMENT

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
Plate: 409 row: C column: 8
High quality sequence stop: 507.

FEATURES

1..625

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

Location/Qualifiers

1..625

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:688000"

/clone_lib="Soares_NHMPU_S1"

/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"

/lab_host="DH10B"

/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH119W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT

109 a 170 c 149 g 92 t

Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Query Match

Best Local Similarity 52.0%; Score 23.4; DB 4; Length 520;

Matches

30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY

3 tccaagcttgccgacgacgaatgagccgacatg 43

Db 62 TCCATGCGCAGCGGCGAGTCCAGCGCAATGTCGCCGAGATT 22

RESULT

A1514427

625 bp mRNA

EST

16-MAR-1999

LOCUS

LD40932.5

Prime ID Drosophila melanogaster embryo P072 Drosophila
melanogaster cDNA clone LD40932 5prime similar to U13637. Y1
F890004649 PID:9535346 SWISS-PROT:P98163, mRNA sequence.

ACCESSION

A1514427

GI:4418489

EST.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 625)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., BroksteinP., Lewis, S. and Rubin, G.M.
BDGP/HMTI Drosophila EST Project
Unpublished (1997)

AUTHORS

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
Plate: 409 row: C column: 8
High quality sequence stop: 507.

JOURNAL

Unpublished (1997)

COMMENT

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
Plate: 409 row: C column: 8
High quality sequence stop: 507.

FEATURES

1..625

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

Location/Qualifiers

1..625

(http://www.ludwig.org.br/scripts/gethtml2.pl?l=6t2-QV1-BT0260-281
099-023-c01&t3=1999-10-28&t4=1)

Seq primer: puc.18 forward
High quality sequence start: 9
High quality sequence stop: 426.

FEATURES
source
Location/Qualifiers

1..528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0260"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 101 a 180 c 152 g 94 t 1 others
ORIGIN

Query Match 49.3%; Score 22.2; DB 134; Length 528;
Best Local Similarity 77.1%; Pred. No. 86;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 9 ctgcggccgcatcagccgaatcagcgccatt 43
||||| ||||| || ||||| ||||| |||||
db 298 CTGCTGCGCTTTCGCGAGGTCGCGCGCAATT 332

RESULT 7
A0841408 558 bp DNA 27-SEP-1999
LOCUS t136718b shotgun sub-library of BAC clone 16K14 Medicago truncatula
DEFINITION genomic clone 16-K-14-022, DNA sequence.
ACCESSION A0841408
VERSION A0841408.1 GI:5930433
KEYWORDS GSS.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE 1 (bases 1 to 558)
Kim, D., Choi, H., Peng, H., Ellis, L. and Cook, D.R.
BAC survey sequencing of Medicago truncatula (1999c)
Unpublished (1999)
Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F.
Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dooc@ppserver.tamu.edu

Other name: BSC-19-22; date: 6/8/99; Submitted to the Database of Genome Survey Sequences (GSS) on 05/27/99; More information is available at 'http://chrysie.tamu.edu/medicago'.
Seq primer: pUC-C
Class: BAC subclone.
Location/Qualifiers
1..558
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone_lib="16-K-14-022"
/note="Vector: pUC18; BAC survey sequences were obtained from sheared BAC DNA subcloned into the SmaI site of pUC18. The template DNA for sequencing was obtained by PCR using universal primers. Sequencing reactions were primed from the pUC-C primer site (CAGGAACACGCTATGACCATGATTACGA)

BASE COUNT 168 a 76 c 95 g 219 t
ORIGIN

Query Match 49.3%; Score 22.2; DB 163; Length 558;
Best Local Similarity 69.8%; Pred. No. 86;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 3 tccaagcttcgagccgcatcagccgaatcagcgccatt 45
||||| ||||| ||||| ||||| |||||
db 444 TGCAGCTTGTGCTGCACAGCGCTTATAGGGAACATTAT 486

RESULT 8
W06254 266 bp mRNA EST 22-MAY-2000
LOCUS tgesy78d04.r1 t9RH tachyzoite cDNA Toxoplasma gondii cDNA clone
DEFINITION tgesy78d04.r1 5', mRNA sequence.
ACCESSION W06254
VERSION W06254.1 GI:1278977
KEYWORDS EST.
SOURCE Toxoplasma gondii.
ORGANISM Toxoplasma gondii.
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.

REFERENCE 1 (bases 1 to 266)
Hehl, A., Manger, I., Marra, M., Sibley, L.D., Ajlola, J.A., Aslett, M.A., Dietrich, N., Dubuque, T., Hillier, L., Kucaba, T., Wan, K.L., Waterston, R.H. and Boothroyd, J.
WashU-Merck-Stanford-NIH Toxoplasma EST project
Unpublished (1996)
Contact: Marra M

TITLE
JOURNAL
COMMENT
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
David Sibley at toxosteborcin.wustl.edu for further information relating to organism, clone or library availability.
Seq primer: t3
High quality sequence stop: 229.
Location/Qualifiers
1..266
/organism="Toxoplasma gondii"
/strain="RH"
/db_xref="taxon:5811"
/clone="tgesy78d04.r1"
/clone_lib="t9RH tachyzoite cDNA"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda ZAP; Site.1: EcoRI; Site.2: XhoI; Toxoplasma RH strain tachyzoites were grown in human foreskin fibroblast cultures in vitro. The library was constructed by K.L. Wan, Cambridge university. cDNAs were synthesized from polyA RNAs by oligo d(T) priming and directionally cloned into the EcoRI to XhoI sites of the Lambda ZapII vector using the ZAP-cDNA synthesis kit (Stratagene). WARNING: the library contains a small percentage of cDNAs derived from the human host cells."

FEATURES
source
Location/Qualifiers
1..266
/organism="Toxoplasma gondii"
/strain="RH"
/db_xref="taxon:5811"
/clone="tgesy78d04.r1"
/clone_lib="t9RH tachyzoite cDNA"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda ZAP; Site.1: EcoRI; Site.2: XhoI; Toxoplasma RH strain tachyzoites were grown in human foreskin fibroblast cultures in vitro. The library was constructed by K.L. Wan, Cambridge university. cDNAs were synthesized from polyA RNAs by oligo d(T) priming and directionally cloned into the EcoRI to XhoI sites of the Lambda ZapII vector using the ZAP-cDNA synthesis kit (Stratagene). WARNING: the library contains a small percentage of cDNAs derived from the human host cells."

BASE COUNT 36 a 66 c 78 g 83 t 3 others
ORIGIN

Query Match 48.9%; Score 22; DB 146; Length 266;
Best Local Similarity 73.7%; Pred. No. 96;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 1 aatccaagcttcgagccgcatcagccgaatcagcgcc 38
||||| ||||| ||||| ||||| |||||
db 81 AATCCACACTTTCGCGAAGATCGGTGAGCATGCGCC 118

RESULT 9
AA952587/c
LOCUS 267 bp mRNA EST 29-OCT-1998
DEFINITION TENS1771 T. cruzi epimastigote normalized cDNA Library trypanosoma
ACCESSION AA952587
VERSION AA952587
KEYWORDS
SOURCE EST
ORGANISM Trypanosoma cruzi.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE 1 (bases 1 to 267)
AUTHORS Verdun,R.E., Di Paolo,N.C., Urmenyi,T.P., Rondinelli,E., Frasch
A.C.C. and Sanchez,D.O.
TITLE Gene discovery through expressed sequence tag sequencing in
trypanosoma cruzi
JOURNAL Infect. Immun. 66 (11), 5393-5398 (1998)
MEDLINE 99003155
COMMENT Contact: Sanchez D.O.
Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
San Martin)
Av. Gral Paz entre Albarillos y Constituyentes, INTI edificio 24
CP(1650) San Martin, Prov. de BS AS, Argentina
Tel: (54-1)752-9639 or (54-1)752-0021
Fax: (54-1)752-0021 or (54-1)752-9639
Email: dsanchez@inti.gov.ar
Seq primer: 17.
Location/Qualifiers
1..267
/organism="Trypanosoma cruzi"
/strain="Cl-Brenner"
/db_xref="taxon:5693"
/clone="1771"
/clone_11b="T. cruzi epimastigote normalized cDNA library"
/cell_type="epimastigote"
/note="cDNA library constructed with oligo dt primed
epimastigote mRNA and cloned in pT7c318D phagemid with
modified polylinker (PHARMACIA)"
BASE COUNT 66 a 79 c 72 g 46 t 4 others
ORIGIN

Query Match 48.9%; Score 22; DB 14; Length 267;
Best Local Similarity 73.7%; Pred. No. 96;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 6 aagcttcgagccgacatcagcgacatattcgagccgacatt 43
|||||
Db 237 AAGCTTGGCGCGCGACGCGCTGCAGTTGGCGCAGTACT 200

RESULT 10
A1667900/c
LOCUS 313 bp mRNA EST 07-JUL-1999
DEFINITION TENC0871 T. cruzi epimastigote normalised cDNA Library Trypanosoma
ACCESSION A1667900
VERSION A1667900
KEYWORDS
SOURCE EST
ORGANISM Trypanosoma cruzi.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE 1 (bases 1 to 313)
AUTHORS Gonzalez Rey,E., Remisz,E., Delgado Garcia,A. and Gonzalez,A.
TITLE Characterization of ESTs from Trypanosoma cruzi epimastigotes
JOURNAL Unpublished (1998)
COMMENT Contact: Delgado Alberto
Departamento de Biologia Molecular, Lab 303
Instituto de Parasitologia y Biomedicina
Consejo Superior de Investigaciones Cientificas C/ Ventanilla No 11
E-18001, Granada, Spain

RESULT 11
W35559
LOCUS 373 bp mRNA EST 22-MAY-2000
DEFINITION TGESF89b07.r1 TgRH tachyzoite cDNA Toxoplasma gondii cDNA clone
ACCESSION W35559
VERSION W35559.1 GI:1317476
KEYWORDS
SOURCE EST
ORGANISM Toxoplasma gondii.
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
REFERENCE 1 (bases 1 to 373)
AUTHORS Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioke,J.A., Aslett,M.A.,
Dietrich,N., Dubugue,T., Hillier,L., Kucaba,T., Wan,K.L.,
Waterston,R.H. and Boothroyd,J.
TITLE WashU-Merck-Stanford-NIH Toxoplasma EST project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@wustl.wustl.edu
David Sibley at toxostep@orcim.wustl.edu for further information
relating to organism, clone or library availability.
Seq primer: T3
High quality sequence stop: 313.
Location/Qualifiers
1..373
/organism="Toxoplasma gondii"
/strain="RH"
/db_xref="taxon:5811"
/clone="tgzy89b07.r1"
/clone_11b="TgRH tachyzoite cDNA"
/lab_host="X1-Blue MRP"
/note="Vector: Lambda ZAP, site_1: EcoRI, site_2: XhoI;
Toxoplasma RH strain tachyzoites were grown in human
foreskin fibroblast cultures in vitro. The library was
constructed by K.L. Wan, Cambridge University. cDNAs were
synthesized from polyA RNAs by oligo d(T) priming and
directionally cloned into the EcoRI to XhoI sites of the

FEATURES
source
1..313
/organism="Trypanosoma cruzi"
/strain="Cl-Brenner"
/db_xref="taxon:5693"
/clone="n42.r"
/clone_11b="T. cruzi epimastigote normalised cDNA library"
/cell_type="epimastigote"
/note="Site 1: EcoRI; Site 2: NotI; cDNA library
constructed with oligo dt primed epimastigote mRNA and
cloned in pT7c318D phagemid with modified polylinker"
BASE COUNT 72 a 91 c 83 g 67 t
ORIGIN

Query Match 48.9%; Score 22; DB 23; Length 313;
Best Local Similarity 73.7%; Pred. No. 97;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 6 aagcttcgagccgacatcagcgacatattcgagccgacatt 43
|||||
Db 255 AAGCTTGGCGCGCGACGCGCTGCAGTTGGCGCAGTACT 218

Lambda ZapII vector using the Zap-cDNA synthesis kit (Stratagene). WARNING: the library contains a small percentage of cDNAs derived from the human host cells."

BASE COUNT 57 a 88 c 110 g 109 t 9 others

ORIGIN

Query Match 48.9%; Score 22; DB 146; Length 373;
Best Local Similarity 73.7%; Pred. No. 99;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 aatccaagcttcgagccgaatcagccgaatcgcgc 38
||||| ||||| ||||| ||||| ||||| |||||
Db 74 AATCCACACTTGCCGAGATCGGTGACGATGCGGCC 111

RESULT 12
AM185272/c 395 bp mRNA EST 19-NOV-1999
LOCUS
DEFINITION
s889c06.y1 Gm-c1023 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1023-1859 5' similar to TR:P93194 P93194 RECEPTOR-LIKE PROTEIN
KINASE.; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

soybean.
Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 395)
Shoemaker,R., Keim,P., Vodkin,L., Erpellding,J., Corvelli,V., Khanna
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Stepoe,M., Theising,B., Allen,M., Bowers
,R., Rittler,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterson,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 Fax:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Putative full length read
vector to vector length read

FEATURES
source

Location/Qualifiers
1..395
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1023-1859"
/clone_lib="Gm-c1023"
/tissue_type="seed coats of greenhouse grown plants"
/lab_host="DH10B"
/note="Vector: pSPORT1; site_1: SalI; site_2: NotI; This
cDNA library was constructed from mRNA isolated from seed
coats (100-200mg) of greenhouse grown plants. The
library was prepared using the lita Technologies
SuperScript cDNA library construction kit. Complementary
DNA was synthesized from mRNA using a poly (dT) sequence
with a Not I restriction site. Sal I linkers adapters
were ligated to the blunt-ended cDNA fragments followed by
Not I digestion. The cDNA fragments were directionally
cloned into the Not I-Sal I restriction site of the
pSPORT1 vector. The ligated cDNA fragments were
transformed into E.coli Electromax DH10B host cells. This

library was constructed by Dr. Lila Vodkin and Dr. Anu
Khanna."

BASE COUNT 122 a 71 c 82 g 120 t

ORIGIN

Query Match 48.4%; Score 21.8; DB 87; Length 395;
Best Local Similarity 70.7%; Pred. No. 1,2e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2 atccaagcttcgagccgaatcagccgaatcgcgc 42
||||| ||||| ||||| ||||| ||||| |||||
Db 247 ATCCAGCTTGTCGACAGATAGAGAGAAATGCTTCTCAT 207

RESULT 13
AV624476/c 436 bp mRNA EST 11-OCT-2000
LOCUS
DEFINITION
reinhardtii cDNA clone LC078c03_r 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizuka@kazusa.or.jp URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers
1..436
/organism="Chlamydomonas reinhardtii"
/strain="Cg"
/db_xref="taxon:3055"
/clone_lib="LC078c03_r"
/clone="Vector: pBluescriptII SK-; site_1: EcoRI; site_2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%
5% to 0.04%
Location/Qualifiers

BASE COUNT 69 a 157 c 116 g 94 t

ORIGIN

Query Match 48.4%; Score 21.8; DB 37; Length 436;
Best Local Similarity 78.8%; Pred. No. 1,2e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 7 agcttcgagccgaatcagccgaatcgcgc 39
||||| ||||| ||||| ||||| ||||| |||||
Db 434 AGCTGGCGGCCCATCATCGGATGCGGCCG 402

RESULT 14

LOCUS

DEFINITION
MGA0062: MGA Library Mycosphaerella graminicola cDNA clone MGA0062
3' similar to OMEGA 6 PATTY ACID DESATURASE, ENDOPLASMIC RETICULUM
ISOZYME 2, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

AM181098 610 bp mRNA EST 17-NOV-1999
AM181098
AM181098
GI:6448332
Mycosphaerella graminicola.

Mon Mar 5 13:21:07 2001

us-09-101-423b-7.rst

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 10:30:25 ; Search time 1021.9 seconds
(without alignments)
225.363 Million cell updates/sec

Title: US-09-101-423B-7
Perfect score: 45
Sequence: 1 aatcaagctgcgcgcgat.....cgaatatgcgcgcattat 45

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues
Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_bal: *
2: gb_bal: *
3: gb_bal: *
4: gb_bal: *
5: gb_bal: *
6: gb_bal: *
7: gb_bal: *
8: gb_bal: *
9: gb_bal: *
10: gb_bal: *
11: gb_bal: *
12: gb_bal: *
13: gb_bal: *
14: gb_bal: *
15: gb_bal: *
16: gb_bal: *
17: gb_bal: *
18: gb_bal: *
19: gb_bal: *
20: gb_bal: *
21: gb_bal: *
22: gb_bal: *
23: gb_bal: *
24: gb_bal: *
25: gb_bal: *
26: gb_bal: *
27: gb_bal: *
28: gb_bal: *
29: gb_bal: *
30: gb_bal: *
31: gb_bal: *
32: gb_bal: *
33: gb_bal: *
34: gb_bal: *
35: gb_bal: *
36: gb_bal: *
37: gb_bal: *
38: gb_bal: *
39: gb_bal: *
40: gb_bal: *
41: gb_bal: *
42: gb_bal: *
43: gb_bal: *

44: em_htg8: *
45: em_htg9: *
46: em_htg10: *
47: em_htg11: *
48: em_htg12: *
49: em_htg13: *
50: em_htg14: *
51: em_htg15: *
52: em_htg16: *
53: em_htg17: *
54: em_htg18: *
55: em_htg19: *
56: em_htg20: *
57: em_htg21: *
58: em_htg22: *
59: em_htg23: *
60: em_htg24: *
61: em_htg25: *
62: em_htg26: *
63: em_htg27: *
64: em_htg28: *
65: em_htg29: *
66: em_htg30: *
67: em_htg31: *
68: em_htg32: *
69: em_htg33: *
70: em_htg34: *
71: em_htg35: *
72: em_htg36: *
73: em_htg37: *
74: em_htg38: *
75: em_htg39: *
76: em_htg40: *
77: em_htg41: *
78: em_htg42: *
79: em_htg43: *
80: em_htg44: *
81: em_htg45: *
82: em_htg46: *
83: em_htg47: *
84: em_htg48: *
85: em_htg49: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24.8	55.1	86896	2	RCU57682
2	23.6	52.4	210047	56	AC010906
3	23.4	52.0	177241	52	HS402611
4	23.2	51.6	6254	32	DMU13637
5	23.2	51.6	14724	58	AC015103
6	23.2	51.6	218971	56	AC011697
7	23.2	51.6	308373	30	AE003495
8	23	51.1	10908	1	AE002294
9	22.2	49.3	2013	2	BAC007ABC
10	22.2	49.3	2565	1	AF030293
11	22.2	49.3	2565	1	RRAJ2069
12	22.2	49.3	10121	1	AE001932
13	22.2	49.3	18622	1	AF012532
14	22.2	49.3	213190	2	BSUB0004
15	22	48.9	11936	1	AE004219
16	22	48.9	66441	32	PEM41P4
17	22	48.9	326301	2	NMA6Z2491
18	21.8	48.4	11627	1	AE004582
19	21.8	48.4	25458	29	SCE22
20	21.6	48.0	47811	59	AC017664
21	21.6	48.0	119526	54	AC007468

misc_feature	1.1569	/organism="Homo sapiens"
misc_feature	/note="assembly_name:Contig13"	/db_xref="taxon:9606"
misc_feature	1670.5999	/chromosome="2"
misc_feature	/note="assembly_name:Contig14"	
misc_feature	6100.12339	
misc_feature	/note="assembly_name:Contig15"	
misc_feature	12440.20468	
misc_feature	/note="assembly_name:Contig16"	
misc_feature	20569.30238	
misc_feature	/note="assembly_name:Contig17"	
misc_feature	30339.44105	
misc_feature	/note="assembly_name:Contig18"	
misc_feature	44206.58754	
misc_feature	/note="assembly_name:Contig19"	
misc_feature	58855.73429	
misc_feature	/note="assembly_name:Contig20"	
misc_feature	73350.94651	
misc_feature	/note="assembly_name:Contig21"	
misc_feature	94752.119515	
misc_feature	/note="assembly_name:Contig22"	
misc_feature	clone.end:SP6	
misc_feature	vector_side:right"	
misc_feature	119616.154482	
misc_feature	/note="assembly_name:Contig23"	
misc_feature	154583.210047	
misc_feature	/note="assembly_name:Contig24"	
misc_feature	50410 c 48661 g 53321 t	
misc_feature	1102	others

	Query Match	52.4%	Score 23.6:	DB 56:	Length 210047:
	Best Local Similarity	76.3%	Pred.	No.1.le=02:	
	Matches 29: Conservative	0:	Mismatches 9:	Indels 0:	Gaps 0:
Oy	1 atccaagcttgcgacgatccaggccgaatatgcgcc	38			
Db_102087	AATCAACACTTTTCAACCAGCGGTGAAAAGAATGCC	102050			

RESULT	3	
LOCUS	HS402G11/c	
DEFINITION	HS402G11 177241 bp DNA PRI 12-DEC-1999 Human DNA sequence from clone RP3-402G11 on chromosome 22q13.31-13.33 Contains the MAPK12 gene for mitogen activated protein kinase 12 (SAPK3), the MAPK1 gene for mitogen activated protein kinase 11 (PRKM1), gene KIAA0315, the gene for a novel protein similar to KIAA0901 and mouse histone deacetylase MHDAA, the gene for a novel protein similar to Xenopus gamma-tubulin interacting protein (yeast SPC98 homolog), the gene for a novel protein similar to yeast and bacterial predicted proteins, the gene for a novel protein similar to C. elegans P3845.2, the gene for a novel protein similar to MRSL and the gene for a novel protein similar to mouse MOV10 (GBL10) and yeast and plant predicted proteins. Contains ESTs, GSSs and fifteen putative CpG islands, complete sequence.	
ACCESSION	AL022328	
VERSION	AL022328.21 GI:5263010	
KEYWORDS	HMG; CpG island; GB110; KIAA0315; KIAA0901; kinase; MAPK11; MAPK12; MHDAA; mitogen activated protein kinase; mitogen-activated; MOV10; MRSL; PRKM1; SAPK3; SPC98. human.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 177241) Coville,G. Direct Submission Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone	
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		

COMMENT

COMMENT

requests: clonerequests@sanger.ac.uk

On Jun 29, 1999 this sequence version replaced gi:5262835.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further Information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information <http://www.wormpep database> can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep

RP3-402G11 is from the library RPCI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RP3-402G11. The true right end of clone RP5-89B14 is at 23403 in this sequence. The true right end of clone RP11-232E17 is at 56734 in this sequence. The true right end of clone RP4-600024 is at 96013 in this sequence.

Location:Qualifiers

1..177241

SOURCE

FEATURES

```

repeat_region      135. .447
                    /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region      657. .959
                    /note="AluSg repeat: matches 1. .310 of consensus"
repeat_region      1373. .1659
                    /note="AluSx repeat: matches 1. .298 of consensus"
repeat_region      1669. .1966
                    /note="AluSg repeat: matches 1. .306 of consensus"
repeat_region      2504. .2574
                    /note="MER58 repeat: matches 170. .243 of consensus"
repeat_region      2663. .2774
                    /note="L2 repeat: matches 2643. .2747 of consensus"
repeat_region      2775. .3072
                    /note="AluSg repeat: matches 1. .300 of consensus"
repeat_region      3073. .3086
                    /note="L2 repeat: matches 2630. .2643 of consensus"
repeat_region      3959. .4455
                    /note="MER1A repeat: matches 31. .527 of consensus"
repeat_region      5044. .5370
                    /note="AluJb repeat: matches 77. .303 of consensus"
repeat_region      6305. .6592
                    /note="AluJo repeat: matches 29. .308 of consensus"
misc_feature       6406. .6907
                    /note="match: GSS: Em:AQ475194"
repeat_region      6853. .7155
                    /note="AluJ repeat: matches 1. .303 of consensus"
repeat_region      7218. .7516
                    /note="AluYb8 repeat: matches 1. .318 of consensus"
repeat_region      9018. .9309
                    /note="AluSx repeat: matches 1. .294 of consensus"
repeat_region      9427. .9542
                    /note="AluJo repeat: matches 30. .150 of consensus"
repeat_region      9543. .9847
                    /note="AluJ repeat: matches 1. .303 of consensus"

```


TITLE
JOURNAL

COMMENT

Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleab, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snil, E., Switskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.
Submitted (11-OCT-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 15, 2000 this sequence version replaced q1:6838484.
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdgpefruitfly@berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently consists of 125 contigs. The true order of the pieces is arbitrary. Gaps between the contigs in this sequence record is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 446: contig of 446 bp in length
* 447 525: gap of unknown length
* 527 1035: contig of 529 bp in length
* 1056 1135: gap of unknown length
* 1136 1828: contig of 693 bp in length
* 1829 1908: gap of unknown length
* 1909 2831: contig of 923 bp in length
* 2832 2911: gap of unknown length
* 2912 3946: contig of 1035 bp in length
* 3947 4026: gap of unknown length
* 4027 4843: contig of 817 bp in length
* 4844 4923: gap of unknown length
* 4924 5371: contig of 448 bp in length
* 5372 5451: gap of unknown length
* 5452 5985: contig of 534 bp in length
* 5986 6065: gap of unknown length
* 6066 6592: contig of 527 bp in length
* 6593 6672: gap of unknown length
* 6673 7486: contig of 814 bp in length
* 7487 7566: gap of unknown length
* 7567 7942: contig of 376 bp in length
* 7943 8022: gap of unknown length
* 8023 8473: contig of 451 bp in length
* 8474 8553: gap of unknown length
* 8554 9234: contig of 681 bp in length
* 9235 9314: gap of unknown length
* 9315 9930: contig of 616 bp in length
* 9931 10010: gap of unknown length
* 10011 10680: contig of 670 bp in length
* 10681 10760: gap of unknown length
* 10761 11304: contig of 544 bp in length
* 11305 11385: gap of unknown length
* 11385 12077: contig of 693 bp in length
* 12078 12157: gap of unknown length
* 12158 12741: contig of 584 bp in length
* 12742 12821: gap of unknown length
* 12822 14882: contig of 2061 bp in length
* 14883 14962: gap of unknown length
* 14963 15971: contig of 1009 bp in length
* 15972 16051: gap of unknown length
* 16052 17440: contig of 1389 bp in length
* 17441 17520: gap of unknown length
* 17521 18329: contig of 809 bp in length
* 18330 18409: gap of unknown length
* 18410 19820: contig of 1411 bp in length
* 19821 19900: gap of unknown length
* 19901 21100: contig of 1200 bp in length
* 21101 21180: gap of unknown length
* 21181 22181: contig of 1001 bp in length
* 22182 22261: gap of unknown length

22262 23685: contig of 1424 bp in length
* 23686 23765: gap of unknown length
* 23766 25198: contig of 1433 bp in length
* 25199 25279: gap of unknown length
* 25279 27114: contig of 1836 bp in length
* 27115 27194: gap of unknown length
* 27195 28549: contig of 1355 bp in length
* 28550 28629: gap of unknown length
* 28630 30577: contig of 1948 bp in length
* 30578 30657: gap of unknown length
* 30658 32556: contig of 1899 bp in length
* 32557 32637: gap of unknown length
* 32637 33822: contig of 1186 bp in length
* 33823 33902: gap of unknown length
* 33903 36202: contig of 2300 bp in length
* 36203 36282: gap of unknown length
* 36283 37948: contig of 1666 bp in length
* 37949 38028: gap of unknown length
* 38029 40445: contig of 2417 bp in length
* 40446 40525: gap of unknown length
* 40526 42921: contig of 2396 bp in length
* 42922 43001: gap of unknown length
* 43002 44588: contig of 2457 bp in length
* 44589 45538: gap of unknown length
* 45539 47806: contig of 2268 bp in length
* 47807 47886: gap of unknown length
* 47887 50521: contig of 2635 bp in length
* 50522 50601: gap of unknown length
* 50602 54569: contig of 3968 bp in length
* 54570 54649: gap of unknown length
* 54650 56887: contig of 2238 bp in length
* 56888 56967: gap of unknown length
* 56968 59805: contig of 2838 bp in length
* 59806 59885: gap of unknown length
* 59886 63021: contig of 3136 bp in length
* 63022 63101: gap of unknown length
* 63102 66948: contig of 3847 bp in length
* 66949 67028: gap of unknown length
* 67029 69188: contig of 2160 bp in length
* 69189 69268: gap of unknown length
* 69269 74338: contig of 5070 bp in length
* 74339 74418: gap of unknown length
* 74419 80192: contig of 5774 bp in length
* 80193 80272: gap of unknown length
* 80273 85047: contig of 4775 bp in length
* 85048 85127: gap of unknown length
* 85128 91343: contig of 6216 bp in length
* 91344 91423: gap of unknown length
* 91424 97193: contig of 5770 bp in length
* 97194 97273: gap of unknown length
* 97274 110572: contig of 13299 bp in length
* 110573 110652: gap of unknown length
* 110653 124105: contig of 13453 bp in length
* 124106 124185: gap of unknown length
* 124186 135484: contig of 11299 bp in length
* 135485 135564: gap of unknown length
* 135565 155302: contig of 19738 bp in length
* 155303 155382: gap of unknown length
* 155383 171964: contig of 16582 bp in length
* 171965 172044: gap of unknown length
* 172045 172568: contig of 524 bp in length
* 172569 172648: gap of unknown length
* 172649 173373: contig of 725 bp in length
* 173374 173453: gap of unknown length
* 173454 173951: contig of 498 bp in length
* 173952 174031: gap of unknown length
* 174032 174486: contig of 455 bp in length
* 174487 174566: gap of unknown length
* 174567 175636: contig of 1070 bp in length
* 175637 175716: gap of unknown length
* 175717 176341: contig of 625 bp in length
* 176342 176421: gap of unknown length
* 176422 176848: contig of 427 bp in length

RESULT 12				
AE001932	AE001932	10121 bp	DNA	BCT 22-NOV-1990
LOCUS	Deinococcus radiodurans	R1 section 69 of 229 of the complete		
DEFINITION	Chromosome 1.			
ACCESSION	AE001932	AE000513		
VERSION	AE001932.1	GI:6458481		

SOURCE ORGANISM	REFERENCE	AUTHORS	TITLE
<i>Deinococcus radiodurans</i>	1 (bases 1 to 10121)	White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.J., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffatt, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Makarova, K.S., Aravind, L., Uitterback, T., Zalewski, C., Vamathevan, J.J., Lam, P., McDonald, L., Dally, M.J., Fraser, C.M., et al.	Genome sequence of the radioresistant bacterium <i>Deinococcus radiodurans</i> RI
<i>Deinococcus radiodurans</i>	2 (bases 1 to 10121)	White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.J., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffatt, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Makarova, K.S., Aravind, L., Uitterback, T., Zalewski, C., Vamathevan, J.J., Lam, P., McDonald, L., Dally, M.J., Minton, K.W., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Fraser, C.M.	Direct Submission
	Submitted (08-Nov-1999)		

FEATURES	
source	GenBank Accession: DQ769811.1 Location: Nucleotide 1..10121 /organism="Deinococcus radiodurans" /strain="R1" /db_xref="taxon:1299" /chromosome="1" 172..1347 /gene="DR0770" 172..1347 /gene="DR0770" note="Similar to GB:U12892 PID:530856 percent identity: 53.56; identified by sequence similarity; putative" /codon_start=1 transl_table=1 product="fimbrial assembly protein PilM, putative" protein_id="AAFI0349.1" /db_xref="GI:6458482" translation="MTSLINFRSRSAIGVEIGTSTIKVVALKAGAPSIQHAWY TPISGRGIVLEPQAVASSELKLAEHRTTRTHATVANCVAYVRNIMVRMDRKI LQAGEPRPVDLKSFAALRALRGMLGELHTKSTLTGTNTGEVALVMETIGASSS INLVKSGDILMTLRNIINSADDEFTALQKAEDVKLGATATPTDDEEDLL NEFDSREGSPARVEEVVPGLDILITEIRSLSEFVROSGDVVIDRFLAGGAKLI GLAAILGALGEGVEVASPMITVTQTDAQGVDTGYLDTMNELTVPLGLALRGVMRG 1340..2056 /gene="DR0771" 1340..2056 /gene="DR0771" note="Identified by Glimmer2; putative" /codon_start=1

gene

```

Product="hypothetical protein"
Protein_id="AA10354.1"
db_xref="Gen:458487"
translation="MVEINILPEYRHRQSPVSWKYASWAAGLIAVYLGCFLLSVSO
LTIQIRRESNAALQOQIDAVAPQKRENDILTAQGELEERTQVATQIRQDRTKQGL
SFEVEPEKNSAASVWNNSTAPAGCEPDLTAAGPEPRTOLDITGSASQCAITGFNLA
ETSDNFQGVDEGQHDHNDIITYETASIGVGGDPSAAGCAPPTAAGATATPAAALAA
AAPATATPAAAPAAAGGAAO"
2053: 2712
(gene="Dr0772")

```


MEDLINE
95286532
REFERENCE
2 (bases 1 to 18622)
AUTHORS
Boriss, R. and Schroeter R.
TITLE
The 55-58 degree segment of the *Bacillus subtilis* chromosome, a
region spanning from the *purA* gene cluster to the *colt* operon
JOURNAL
3 (bases 1 to 18622)
AUTHORS
Boriss, R.
TITLE
Direct Submission
JOURNAL
Submitted (03-JUL-1997) Boriss R., Humboldt University, Institute
of Biology, Chausseestrasse 117, Berlin, Germany, D-10115
FEATURES
Location/Qualifiers
1. .18622
/organism="Bacillus subtilis"
/strain="168 trpc2"
/db_xref="taxon:1423"
/map_55 to 58 degree"
/map_55 to 58 degree"
/gene="yefA"
/gene="yefA"
/gene="yefA"
/note="Similar to E. coli hypothetical RNA
methyltransferase: SwissProt Accession Number P55135;
Similar to *Chlamydia trachomatis* Hom1: SwissProt Accession
Number P55137"
/codon_start=1
/citation=[2]
/transl_table=11
/product="yefA"
/protein_id="AAB66480.1"
/db_xref="GI:2318067"
/translation="AKRLRSPADNCGDEAGSYCVL"
complement(227. .1140)
/gene="yefB"
complement(227. .1129)
/gene="yefB"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefB"
/protein_id="AAB66472.1"
/db_xref="GI:2318059"
/translation="MYNEGVTSPQLARKKATDKWHSIRKILENPHYTGLOOCR
EYPSVTSKRRSVNLENOIIEKSHPIPLDEPLIYOEILITRRRRPQAVHLEF
NTAVCKDCRSRHHKMKRKGVCYSYNGHGLKACSDHYSEIDLTPKVTINLKITYLK
ESKENYKEISEKALKYKEKEIKIEINKLNDKRDKNLVTISLANGVTSKEDYOL
AINTNEDISNFEITLHLSKIDLEFORIEKEIIDFKKSIDKEKESGLTPEMLHLVD
EIEVHANGTIEINVRPREPTVPSA"
1024. .1445
/gene="yefC"
/gene="yefC"
1024. .1030
/gene="yefC"
1068. .1445
/gene="yefC"
/note="similar to Enterococcus faecalis plasmid PY117 orf4
encoded by GenBank Accession Number D78257"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefC"
/protein_id="AAB66473.1"
/db_xref="GI:2318060"
/translation="MSVAFPLRASCGLVTPSLYNPLATASKRPPAIYSKILITIF
SAFAEPRDMIVERTQEGKMLAKONDFRGRKRKTKQOINAHLLLENHSTKQYED
MTGISVSTIVRAKKAKAAEAING"
complement(1133. .1140)
/gene="yefB"
1505. .4159
/gene="yefB"
1505. .1512
/gene="yefA"
1520. .4159
/gene="yefA"
/gene="yefA"

/note="similar to Methanococcus janaschii plasmid PUR801
hypothetical protein MJEC502 encoded by GenBank Accession
Number U77119"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefA"
/protein_id="AAB66474.1"
/db_xref="GI:2318061"
/translation="MALIDLEDKIAEIVNREDSDFELLEGVYDVPRATTIRLKG
ONLTKRVGVHLLKRWYFEKARKGLFDALIDIEQVEYLKAPRYLVTVYDYLAK
DKITLALDVFKEELPOYDFPLAWKGIKVEKEENPADIKAAERFADYVLEKEN
NIETNRGLDFILRLFCFPAEDNDIEKRNSETULIKITLFEEDSNLNKILPADLPIV
LDKNERDVPYSLKEPRYNGOLFTEPRHIELEFSKSKRLITECEGLLNWAKINDIF
GSMIOAVASESRSTLGHMTYSVPNIMKATKFLDKLNOSTLDVDTYKLENTLR
ICKIKFEPDAGSGNELLITVYELRMEINILIKRQELLGELYLVSTLSQPGIEI
EDFAHDVAKLSLMAIEHOMNEBELKNEVNAVAPTLPLHAGIRGANAIRMEVETCP
AAGSEEVYVGNPYPYLGSKKONKERSDMLSTFGKVKNGKMDYISAMFYCAKAST
TNAKVAFVSTNSVTOGEQVSIIMNELFKFGIINFNKSNFKANNAKNAVIVYVIG
FGPIDKTVKXYLFVDETRKLVNISPIYLDGNNIIVSSRTKISLPRLHGNMNDG
GGLITITTYTALINKYPLVYPERKFTGSVEFINGGLRYCLMNEAKYETKSNPLI
QERISIKNHRKESDQGTNKLALTPMKERDTHETJNSIIVPSVSENRYIEMGLA
GADTILSNLIVYDAEIVLIGILMSRMHMTWKAAGRLKTDYRSAGLCYNEFPPI
ELSTRKNEIEEAILIIDLREOGGTIAEIVNPSTMPTELKVAHEKLDGIVERAYRQ
KQESDEERLEVLKLYQEMTER"
4152. .4161
/gene="yefB"
4152. .5967
/gene="yefB"
4171. .5967
/gene="yefB"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefB"
/protein_id="AAB66475.1"
/db_xref="GI:2318062"
/translation="MENITEIYNOTGSKKTNEYGMRENQARAREKRNQOYLTKAP
PASGSRALMFIGLKLINOGLRKYAVAPERSITSSRQNDKATGCFEMKWDPRN
NLTVGDNKSVKSFVFMESDQVILICHTSTIRFAFEKIDDAFNCILAIIDEHHVS
ADVNSRLGELRSIIHNSSAHIVAMTGSYFRSDSVILPEDEELFDVYTSYEOLD
GYEYLKGFIGIYHGYQTSAINEVLDNTKTYIHI PNVSSESTKQYDEVGKILD
MGEVYEGODDDGIIVYKRHSQGTILADIADVDOGVRENVAIYARNTEALDDIIL
ALGMKEGDDMPCEHTLVYGRSISTETIYOTIGCTPDSVAKYCAOFTNLIIMDDAK
DEVYITVYTMKALISGLMKOVLTPPDKERKRNSEKSTTGLVKGLEKSTDE
NKKIILENDINDLAKTCKIHSKTFSEGVDPKVLNKLIPKIVQKVPNLSDEIEE
VQOHVLDIVMKAQSEVYKSEFTIRMDKFPINIEDLDINLIDSINPOKAFEVLSKE
LNSPYLKLIOETIDAKRINFDEDELVFTMPKV"
6064. .6069
/gene="yefC"
6064. .7228
/gene="yefC"
6080. .7228
/gene="yefC"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefC"
/protein_id="AAB66476.1"
/db_xref="GI:2318063"
/translation="MASNRYHSINFINESQLEFYOITPPPKTKOKAOVPEVEKPLEIT
EYVKEGREGPOKVPIDIFERSLASRLIGIRKPDHMEYLKEDEIGILLEVRSKELTIP
KISSIDIDILKSLSDLLDNLNNDITRSIFDTSIQKVTMPREYAKRKKIKDKRFE
ELFKOHRKITEGKRIILFKNEODIQSFFILKGVLLIYVEDGERRKAKGTNARL
RCIFENGESDMLRSLSAELVYKRGARDNYEDTLDDVNRDSDVSTGFIYVLSLSD
POISSIKNLKYGITFTGSVEYRIRNAENQSTYLVAPVITVYGVFNNAKSKFETAIH
HALENNDLVSIIIGANGKMLVPKEFWVTLEDLQAVIDEIVAMVHLVD"
complement(7490. .8180)
/gene="yefB"
complement(7490. .7795)
/gene="yefB"
/citation=[2]
/codon_start=1

/db_xref="SPTRMBL:P96699"
/translation="MPYOINANIKINGVQNGNIDVGPVHNSHTANSKYFGANFSL
GDLSPSSLNTGNTGSDVDSDQIGNPSAPISNQI"
complement(1455..1727)
/gene="ydgB"
complement(1455..1727)
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12376.1"
/db_xref="GI:2632870"
/db_xref="SPTRMBL:P96700"
/translation="MPSTVINLYIKINSISGNGSTIGFAVNSPTNNKSGGTNS
FGDRSPESYMENTLNDPDVNDQTSIGNSDPSNNAPIPPILD"
2026..2613
/gene="ydgC"
2026..2613
/gene="ydgC"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12377.1"
/db_xref="GI:2632871"
/db_xref="SPTRMBL:P96701"
/translation="MKSQVSESRRLKKAANESVGRFDAYSEIVKAGTQPSF
KLYRQSEAFPAELITDPHSRVRKRLTSLLENGLNTEDVSKRLVLAVERFPLDD
KDLTKGFPLNPEAKOKKDLAMVLEKLEAQRLEQFHSLEDMETVAECLVGMIEHL
JESFLNGIKDPSLAEEVNVLLIYGMIPKGNQVR"
2610..2954
/gene="ydgD"
2610..2954
/gene="ydgD"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12378.1"
/db_xref="GI:2632872"
/db_xref="SPTRMBL:P96702"
/translation="MISIMMKVSLAVMLAGGIKVSFVPEVHEMRHXYQPLFLTV
TGLEIGALAMTGAIGWNRVYAAIGAGVLFVVLAMAGIAHMFRAQGVIMAIQMICL
IVSIMIMSGYT"
3074..3091
/gene="ydgD"
3117..3590
/gene="ydgE"
3117..3590
/gene="ydgE"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12379.1"
/db_xref="GI:2632873"
/db_xref="SPTRMBL:P96703"
/translation="MAFTLEDTEEEFEARGSVONFAQONTISGWTKEAFKESK
QATENMIPNGRDSNNHFNMTNDQGRKMWLMYADPLHPQREAFIYSFGILEARG
KGLAQALQTLDERARLGAERLALHFAHNETAVVLYOKMGYAMTNRIRKOLC"
complement(3705..3722)
/gene="exp2"
complement(3705..3722)
/gene="exp2"
complement(3705..3722)
/gene="exp2"
complement(3750..5393)
/gene="exp2"
/codon_start=1
/transl_table=11
/product="ATP-binding transport protein"
/protein_id="CAB12380.1"
/db_xref="GI:2632874"
/db_xref="SWISS-PROT:P39115"
/translation="MKEIVTLTVNSYEVKQGVFKHVNASVQGGDIIGIKGNAGKS

Query Match 49.3%; Score 22.2; DB 2; Length 213190;
Best Local Similarity 69.8%; Pred. No. 3.5e-02;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 3 tccaagcttgccgcagcagcgaatgcgcgacatcat 45
Db 154535 TCAGAGTTTCGGGTGCTCATGCTATGACCATTTT 154577
RESULT 15
AE004219
LOCUS AE004219 11936 bp DNA BCT 31-JUL-2000
DEFINITION Vibrio cholerae chromosome I, section 127 of 251 of the complete
chromosome.
ACCESSION AE004219 AE003852
VERSION AE004219.1 GI:9655899
KEYWORDS
SOURCE
ORGANISM
Vibrio cholerae.
Vibrio cholerae
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 11936)
AUTHORS Heidelberg, J.F., Eissen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Emmolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Uitterback, T., Fleischmann, R.D.,
Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
TITLE DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
JOURNAL Nature 406, 477-483 (2000)
REFERENCE 2 (bases 1 to 11936)
AUTHORS Heidelberg, J.F., Eissen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Emmolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Uitterback, T., Fleischmann, R.D.,
Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

FEATURES		Location/Qualifiers
source		1. 11936
gene		/organism="Vibrio cholerae"
CDS		/strain="N16961"
gene		/serotype="O1"
CDS		/db_xref="taxon:666"
gene		/chromosome="1"
CDS		/note="biotype: El Tor"
gene		/complement(98. .220)
CDS		/gene="VC1404"
gene		/complement(98. .220)
CDS		/gene="VC1404"
gene		/note="identified by Glimmer2; putative"
CDS		/codon_start=1
gene		/transl_table=11
CDS		/product="hypothetical protein"
gene		/protein_id="AAF94561.1"
CDS		/db_xref="GI:9655900"
gene		/translation="MENLEIQOTLIKISTKNIDADLAQILFPLNKNPCF"
CDS		460. .2247
gene		/gene="VC1405"
CDS		460. .2247
gene		/gene="VC1405"
CDS		/note="similar to GB:X61367 PID:48195; identified by sequence similarity; putative"
gene		/codon_start=1
CDS		/transl_table=11
gene		/product="methyl-accepting chemotaxis protein"
CDS		/protein_id="AAF94562.1"
gene		/db_xref="GI:9655901"
gene		/translation="MSIYLKALIPSRHRHFWSHILHVTWVNRVTKMFWNNLSITOK MAIVGCIALLTAVSTISITMKRVAVEIEMAHNLPLAKMTMTVHOEGATTL EVPFRAGTIESSGQKQNOYOQALMEYLNKRKQENOSROLLSEIETISISY LANLEDLDSREHEERKLNELNELSGREKVLVADABQILEQDLSIDRLID ILRKLEQATQSVLVEDEERBALMGWVLSAFLLEGLIGFAFSKQVMAIARAK LANEMAEFNFSRAKVTGDEIGLITSMNTAQSLSHVGEVDIRANTYASTVOLA SAESNKSQVQOQANTQVTSAMAQMTITEVASSESSAATRAQENAYSCDV LKTETVSSOLVANAQSOQMIVELEASTROIESFVLVEGISECTNLALNAIEAA RGEQGRGPAVVADEVRALASRSQATHEIKGILTIVERRASATMTKIDSSPOIEES FESSTRAKOLDISINALELITSANTOVAASBESVAADEISHNTDIRDAGETIML SAQETPAQSEELAQOAGLKLMPVIVS"
gene		2326. .3642
CDS		/gene="VC1406"
gene		2326. .3642
CDS		/gene="VC1406"
gene		/note="similar to GB:X61367 PID:48195; identified by sequence similarity; putative"
gene		/codon_start=1
CDS		/transl_table=11
gene		/product="methyl-accepting chemotaxis protein"
CDS		/protein_id="AAF94563.1"
gene		/db_xref="GI:9655902"
gene		/translation="MIMKPFMPWPKAAAOSEVEQHNHDSIVSALKRNLAVTEPQ GILDLNALFLSAMGYTKKEVIGQHKIFCDPETVNSLEYAOFKSLAOGAQRKMF RKDKGSSIWIEATYMPVRNRAQDVYKIVKADYKARKITADKQAVSHSDRSSAM IQENDGTIKENENMLKATGRLEIEVGHMHMFCDDKREKPNFEMELIAGSEKS GPHRRTRGQDMLLEATYNPINQAGVTVQVKKFASDVTDOVLKAAHKEASOMAO TSAETVAVASGREMLDAATITAGSTTESIACANALMTDLSQSQRTIQTIVTINKIA EQTNLALNAIEAARAGEYGRGPAVVADEVRSLASNTSOATSEIDNIYKRSNLTEQ SGQTMQIQAKTTEFNHMDQOTSLIEQIOHAENVOKTVSNIVDE"
gene		3824. .5017
CDS		/gene="VC1407"
gene		3824. .5017
CDS		/gene="VC1407"
gene		/note="similar to GP:328677; identified by sequence similarity; putative"
gene		/codon_start=1
CDS		/transl_table=11
gene		/product="ATP-dependent RNA helicase RhlE"
CDS		/protein_id="AAF94564.1"
gene		/db_xref="GI:9655903"
gene		/translation="MSFSQGLSDVLVQVVAQLYGQKPTHTIOGAIIPVILQGRDLIAA AQTGKTASFVLPILKIQGQTORKKRVRALIVPTRELAMQVAEKQYKGDITGL

gene		complement(5076. .5678)
CDS		/gene="VC1408"
gene		complement(5076. .5678)
CDS		/gene="VC1408"
gene		/note="similar to GB:M97208 SP:P32398 PID:143046 PID:2226233 GB:AL009126; identified by sequence similarity; putative"
gene		/codon_start=1
CDS		/transl_table=11
gene		/product="transcriptional regulator, TetR family"
CDS		/protein_id="AAF94565.1"
gene		/db_xref="GI:9655904"
gene		/translation="MRVSEERKQALILEIAKDSFTKQGFQOTSMISHIAKMGSKATL YNFSSKERIFPAVMESSATFOIALSFLSIDHRELBKPLLEEGYFELNSVLTODAMS IYRMATHEADRSATIGRHFYENGKRGMARLSRYITQOICGSLKECDPMIAMHFKAL LSAQTFEPFSLGAIKPTSELKATTQRAVEAFLITAAD"
gene		5785. .7239
CDS		/gene="VC1409"
gene		5785. .7239
CDS		/gene="VC1409"
gene		/note="similar to GP:3378278; identified by sequence similarity; putative"
gene		/codon_start=1
CDS		/transl_table=11
gene		/product="multidrug resistance protein, putative"
CDS		/protein_id="AAF94566.1"
gene		/db_xref="GI:9655905"
gene		/translation="MKNVQYVGLHKAAPYFASLSAFVLTCGSPHPHYDLATW EAEISSTVFISHQAEIMDPSAMWQRYQDQNLHLEALQSPSLEMAARLKQAO GFARQAGARSPDLGLAASATBESKVSERYQSAAPPDGMNDYGLTLNFPYDEDEWKN RAAVVATSELAAEAESVAARLMISTIANAVAEALRANQETLVAALQVANKVE LEBKRYANGLETIGSVQAKVAASVEALIGIOESIQOKNLAALVQGPFRASII EBPHILITSVYGIPEFAGVGLGRADITVAARRAANAQOVIAQOYRPDYTLISAF IGYQAFGDHLPSCNDACAGIATYLPFTGRLLEGQLTSAPARQOEAQVQVNCNTLY QALYEVADVTSQALQARKNTEQAVQAEQALHAINRYQGLATYIDVLYAEBSL LNNORALVNLQSAFSLDLALHLAGGFETTES"
gene		7220. .8440
CDS		/gene="VC1410"
gene		7220. .8440
CDS		/gene="VC1410"
gene		/note="similar to GP:3378278; identified by sequence similarity; putative"
gene		/codon_start=1
CDS		/transl_table=11
gene		/product="multidrug resistance protein VceA"
CDS		/protein_id="AAF94567.1"
gene		/db_xref="GI:9655906"
gene		/translation="MKQNLNRLSSNNSNNTPEFQVRAQSRKKGFLGLAAIIVA GGSYALVYHPIGSRVISTDNAYAAAEIATVPAVGIIAQNVAVDVEYVQOGVIVOL DDIDARIALYHPIGLADLAKRRVRYSLANDEGLSAVEAOEAQEVKAKQAKQAQDF ERAKIDLSREDDVRSVSGEELTNAKGFEAPQVQNLNARAKMAAOATKSTIGS QRNALTLTQNTYDSNPVILAKAREQAKIDIEPVIYIARAPISGVIARQKRGSTVO VGMPLTVVPTDHTIYDANKEVELDVKQGVYTTADLIGDVTYHGVVAFSSGT GSAFEMIPAQNAQGNMIRKVVQRILRIELDPKQAVPLQVGLSMWATIDTACTTPO TLVQYRAKVSDEQG"
gene		8450. .9985
CDS		/gene="VC1411"
gene		8450. .9985
CDS		/gene="VC1411"
gene		/note="similar to GP:2815578; identified by sequence similarity; putative"
gene		/codon_start=1
CDS		/transl_table=11
gene		/product="multidrug resistance protein VceB"
CDS		/protein_id="AAF94568.1"
gene		/db_xref="GI:9655907"
gene		/translation="MSHNADNEMQPLSGNALPFGALCLAMANFLALIDPTTIANVSVSN

gene

/gene="VC1412"
9939. .10055

```
/note="identified by Glimmer2; putative"  
/codon start=1
```

```

/ transl_table=11

```

```
/product="hypothetical protein"  
/protein_id="AAF94569.1"
```

```
/db_xref="GI:9655908"
```

```

/ciandsidation="MRQNPskRWIPLLfIKRQRNTHGITPLFSRAYEGGD"
complement(10060. .11796)
/cenae="vc1A13"

```

```
complement(10060. .11796)
```

Query Match
Best Local

Best Local Similarity	/3.7%;	Pred. No. 6.2e+02;
Matches	28;	Conservative 0;
		Mismatches 10;
		Indels

Matches	28;	Conservative	0;	Mismatches	10;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

3 tccaagctgcgcccgatcagccgatatgcggcgc 40

D6 6289 TCCGAGCTTGCTGCCGCTGAGGCAGAGTGTGGCTGC 6326

Search completed: March 4, 2001, 12:03:42
Job time: 5597 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 10:32:15 ; Search time 106.76 Seconds
(without alignments)
67,930 Million cell updates/sec

Title: US-09-101-423B-7
Perfect score: 45
Sequence: 1 aatcaagcttgccgcat.....cgaatatggccgcattat 45

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

* Result No.	Score	Query Match	Length	DB ID	Description
C 1	21.2	47.1	3481	3	US-09-136-251-1
C 2	20.4	45.3	41	1	US-08-200-900A-26
C 3	20.4	45.3	41	4	PCT-US94-00616-26
C 4	20.2	44.9	2885	3	US-08-948-705-26
C 5	19.8	44.0	3415	1	US-08-054-077C-1
C 6	19.8	42.2	638	1	US-08-469-667-1
C 7	19.8	42.2	638	4	PCT-US95-07289-1
C 8	18.8	41.8	38506	3	US-09-320-878-19
C 9	18.6	41.3	37	2	US-08-839-581A-21
C 10	18.6	41.3	62	5	5332671-12
C 11	18.6	41.3	1708	4	PCT-US95-02481-3
C 12	18.6	41.3	8051	2	US-08-576-628A-2
C 13	18.4	40.9	88	1	US-08-706-037-16
C 14	18.4	40.9	88	2	US-09-005-397-16
C 15	18.4	40.9	68750	3	US-09-335-409-1
C 16	18.2	40.4	1374	1	US-08-278-630A-9
C 17	18.2	40.4	1424	1	US-08-403-634-3
C 18	18.2	40.4	1424	3	US-08-913-441B-3
C 19	18.2	40.4	1584	3	US-08-907-229-1
C 20	18.2	40.4	1965	1	US-08-258-420-9
C 21	18.2	40.4	2001	3	US-08-850-961-13
C 22	18.2	40.4	2943	1	US-08-042-747A-7
C 23	18.2	40.4	6028	3	US-09-011-745-5
C 24	18.2	40.4	495	1	US-08-406-248-1
C 25	18.2	40.4	1194	4	PCT-US94-09700-10
C 26	18.2	40.4	1721	1	US-08-241-766-3
C 27	18.2	40.4	2100	1	US-08-154-915-5
C 28	18.2	40.4	2106	1	US-07-970-462A-1

Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 10, Appli
Sequence 22, Appli
Sequence 22, Appli
Sequence 22, Appli
Sequence 22, Appli
Sequence 22, Appli
Sequence 8, Appli
Sequence 8, Appli
Sequence 8, Appli

C 29 18 40.0 2106 4 PCT-US92-10904-1
C 30 18 40.0 2106 4 PCT-US94-09700-1
C 31 18 40.0 2121 1 US-08-574-043A-1
C 32 18 40.0 2121 2 US-08-795-015-1
C 33 18 40.0 2121 4 PCT-US94-12936-1
C 34 18 40.0 2121 4 PCT-US96-11886A-1
C 35 18 40.0 2245 2 US-08-617-801A-3
C 36 18 40.0 3051 1 US-08-241-766-10
C 37 18 40.0 6914 1 US-08-920-812-22
C 38 18 40.0 6914 1 US-08-920-812-22
C 39 18 40.0 6914 1 US-08-921-177C-22
C 40 18 40.0 6914 1 US-08-362-577C-22
C 41 18 40.0 6914 2 US-08-920-828-22
C 42 17.8 39.6 981 1 US-08-259-148A-8
C 43 17.8 39.6 981 1 US-08-484-054-8
C 44 17.8 39.6 981 2 US-07-876-941A-8
C 45 17.8 39.6 984 1 US-08-240-049B-8

ALIGNMENTS

RESULT 1
US-09-136-251-1/c
; Sequence 1, Application US/09136251A
; Patent No. 6127156
; GENERAL INFORMATION:
; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: MIYAZAKI, Taro
; APPLICANT: OJIMA, Setsuko
; APPLICANT: SHINJOH, Masako
; APPLICANT: TOMIYAMA, No. 61271561bun1
; TITLE OF INVENTION: D-SORBITOL DEHYDROGENASE GENE
; FILE REFERENCE: D-Sorbitol Dehydrogenase Gene
; CURRENT APPLICATION NUMBER: US/09136,251A
; CURRENT FILING DATE: 1998-08-19
; EARLIER APPLICATION NUMBER: EP 97114432.4
; EARLIER FILING DATE: 1997-08-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3481
; TYPE: DNA
; ORGANISM: Gluconobacter suboxydans
US-09-136-251-1

Query Match 47.1%; Score 21.2; DB 3; Length 3481;
Best Local Similarity 76.5%; Pred. No. 11;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 ccaagcttgccgcatcagccgaatgctgcgc 37
DB 388 CCAAGCGTGGCCCATGACATGAATAGCGCGC 355

RESULT 2
US-08-200-900A-26/c
; Sequence 26, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

us-09-101-423b-7.rni

Mon Mar 5 13:21:06 2001

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinert, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-200-900A-26

Query Match 45.3%; Score 20.4; DB 1; Length 41;
Best Local Similarity 71.1%; Pred. No. 13;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 8 gcttgccgcgacagccgaatatgcggccgattat 45
Db 41 GCTTGGACTTAACGAGCTGACTTGGCGCGCATCT 4

RESULT 3
PCT-US94-00616-26/c
; Sequence 26, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00616
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; PCT-US94-00616-26

Query Match 45.3%; Score 20.4; DB 4; Length 41;
Best Local Similarity 71.1%; Pred. No. 13;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 8 gcttgccgcgacagccgaatatgcggccgattat 45
Db 41 GCTTGGACTTAACGAGCTGACTTGGCGCGCATCT 4

RESULT 4
US-08-948-705-2
; Sequence 2, Application US/08948705A
; Patent No. 6043084
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND METHODS FOR DIAGNOSING AND
; TITLE OF INVENTION: TREATING COLON CANCER
; FILE REFERENCE: LUD-5506-JEL/NDH
; CURRENT APPLICATION NUMBER: US/08/948,705A
; CURRENT FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2885
; TYPE: DNA
; ORGANISM: Homo Sapiens
; US-08-948-705-2

Query Match 44.9%; Score 20.2; DB 3; Length 2885;
Best Local Similarity 75.8%; Pred. No. 25;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 8 gcttgccgcgacagccgaatatgcggccgc 40
Db 2715 ggtgggggcccagacagcccaagctgcggcgc 2747

RESULT 5
US-08-054-077C-1/c
; Sequence 1, Application US/08054077C
; Patent No. 5527679
; GENERAL INFORMATION:
; APPLICANT: HEMLER, MARTIN E.
; APPLICANT: RAMASWAMY, HEMAVATHI
; TITLE OF INVENTION: HUMAN INTEGRIN 5 SUBUNIT PROTEIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/054,077C
; FILING DATE: 27-APR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694314
; FILING DATE: 01-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 40937
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3415 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: mat peptide
```


TELEPHONE: 404-818-3700
TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1708 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US95-02481-3

Query Match 41.3%; Score 18.6; DB 4; length 1708;
Best Local Similarity 65.9%; Pred. No. 95;
Matches 27; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 4 ccaagcttgccgagccgacgaataatgcccgcacat 44
DB 260 CCCTGCTGCACAGCATCATGTCACATCTGTCCTCTTA 220

RESULT 12
US-08-576-626A-2/c
Sequence 2, Application US/08576626A
Patent No. 5998194
GENERAL INFORMATION:
APPLICANT: Summers, R.G.
APPLICANT: Katz, L.
APPLICANT: Donadio, S.
APPLICANT: Staver, M.J.
TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/576,626A
FILING DATE: 21-DEC-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 3857, US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847) 938-3137
TELEFAX: (847) 938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 8051 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-576-626A-2

Query Match 41.3%; Score 18.6; DB 2; length 8051;
Best Local Similarity 72.7%; Pred. No. 1,1e+02;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 11 tgcgagcagatcagccgaataatgcccgcacat 43
DB 6190 TCGGCGGACGACGTCGAGCAGGTGGCGGACT 6158

RESULT 13
US-08-706-037-16

Sequence 16, Application US/08706037
Patent No. 5770419

GENERAL INFORMATION:
APPLICANT: Xu, Feng
APPLICANT: Berka, Randy M.
APPLICANT: Wahleithner, Jill A.
TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH
TITLE OF INVENTION: ENHANCED ACTIVITY
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5770419 of No. 5770419 of No. 5770419 of America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,037

FILING DATE: 30-AUG-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/002,800

FILING DATE: 1-SEP-1996

ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.

REGISTRATION NUMBER: 38,711

REFERENCE/DOCKET NUMBER: 4526, 200-US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 88 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA
US-08-706-037-16

Query Match 40.9%; Score 18.4; DB 1; length 88;
Best Local Similarity 63.6%; Pred. No. 79;
Matches 28; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 aatccagcttgccgagccgaataatgcccgcacat 44
DB 20 AATTCAGCTTGCGACGCTTTAAACCAATTTGAGATCTCTTAA 63

RESULT 14
US-09-005-397-16

Sequence 16, Application US/09005397
Patent No. 5972670

GENERAL INFORMATION:

APPLICANT: Xu, Feng

APPLICANT: Berka, Randy M.

APPLICANT: Wahleithner, Jill A.

TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH
TITLE OF INVENTION: ENHANCED ACTIVITY

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5972670 of No. 5972670 of No. 5972670 of America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York

STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,397
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/706,037
FILING DATE: 30-AUG-1996
APPLICATION NUMBER: US 60/002,800
FILING DATE: 1-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4526,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
-US-09-005-397-16

Query Match 40.9%; Score 18.4; DB 2; length 88;
Best Local Similarity 63.6%; Pred. No. 79;
Matches 28; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 1 aatccaagcttcgagccgcatcagcgccgacatgcgacgcat 44
||| ||||| | | | | | | | | | | | | | | | |
Db 20 AATTCAAGCTTGTGACGCTTTAAACCAAAATTGAGATCTTTAA 63

RESULT 15
US-09-335-409-1
Sequence 1, Application US/09335409
Patent No. 6121029
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLIONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/335,409
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-335-409-1

Db 36360 cggccgagcagtcacactatgcgagcgc 36387
Search completed: March 4, 2001, 12:05:17
Job time: 5582 sec

Query Match 40.9%; Score 18.4; DB 3; length 68750;
Best Local Similarity 78.6%; Pred. No. 1,7e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
OY 13 cggccgacgagcgcgacatgcgagcgc 40
||||| ||| | | ||||| | |

Mon Mar 5 13:21:06 2001

us-09-101-423b-7.rni

Page 7

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 10:32:45 ; Search time 96.53 Seconds
(without alignments)
175.125 Million cell updates/sec

Title: US-09-101-423B-7
Perfect score: 45
Sequence: 1 aatccaagcttgccgcgat.....cgatatgcgcgcattat 45

Scoring table: IDENTITY_NUC
Gapop 10.0, Capext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq_36.*
2: /cgn2_2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
3: /cgn2_2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
4: /cgn2_2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
5: /cgn2_2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
6: /cgn2_2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
7: /cgn2_2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
8: /cgn2_2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
9: /cgn2_2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
10: /cgn2_2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
11: /cgn2_2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
12: /cgn2_2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
13: /cgn2_2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
14: /cgn2_2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
15: /cgn2_2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
16: /cgn2_2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
17: /cgn2_2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
18: /cgn2_2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
19: /cgn2_2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
20: /cgn2_2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
21: /cgn2_2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
22: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	45	18	T72784 DNA tag. Synthetic
2	21.4	47.6	42	18	Primer HUCCN NOR f
3	21.2	47.1	500	20	T. gondii immunoge
4	21.2	47.1	3481	20	DNA sequence of so
5	20.8	46.2	1503	19	Hamster oral papill
6	20.8	46.2	7647	19	Nucleotide sequenc
7	20.8	46.2	8907	21	Stealth virus nucl
8	20.8	46.2	9181	20	Lambda INNER prime
9	20.4	45.3	41	15	Rifin 3193 gene.
10	20.4	45.3	939	21	Sequence encoding
11	20.4	45.3	4260	9	Sequence of Mycoba
12	20.4	45.3	4380	9	N80222

C	13	20.4	45.3	4380	19	V05708
C	14	20.2	44.9	1818	21	Z46842
C	15	20.2	44.9	2885	20	X40055
C	16	20.2	44.9	5398	12	O13284
C	17	20	44.4	2178	20	T12261
C	18	19.8	44.4	3415	17	T36481
C	19	19.8	44.0	7673	19	V58229
C	20	19.6	43.6	3183	20	X26611
C	21	19.4	43.1	524	20	X84148
C	22	19.4	43.1	524	21	Z53380
C	23	19.4	43.1	546	20	Z12160
C	24	19.4	43.1	753	21	Z53377
C	25	19.4	43.1	753	21	Z53374
C	26	19.4	43.1	2352	21	Z53296
C	27	19.2	42.7	619	21	Z53637
C	28	19.2	42.7	834	21	Z53639
C	29	19.2	42.7	1095	21	Z54498
C	30	19.2	42.7	1842	21	Z56349
C	31	19	42.2	417	20	V90226
C	32	19	42.2	638	18	T45880
C	33	19	42.2	638	19	V16668
C	34	19	42.2	1173	20	X91764
C	35	19	42.2	1218	20	X91630
C	36	19	42.2	4394	13	O21604
C	37	19	42.2	5474	16	O90251
C	38	18.8	41.8	2302	21	A15621
C	39	18.8	41.8	2302	21	A15622
C	40	18.8	41.8	2465	18	T58086
C	41	18.8	41.8	2465	21	A15620
C	42	18.8	41.8	4248	21	A27364
C	43	18.8	41.8	7059	15	O55004
C	44	18.8	41.8	7812	20	X12962
C	45	18.8	41.8	13842	21	Z87297

ALIGNMENTS

Mycobacterium tube
Semaphorin K1 gene
Colon cancer assoc
P. dentrificans co
Neisseria gonorrho
Human Integrin bet
Omega-cytohexasane
Trypanosoma cruzi
DNA encoding human
Neisseria gonorrho
Neisseria gonorrho
Neisseria gonorrho
Neisseria meningit
Neisseria meningit
Neisseria meningit
Escherichia coli f
EST clone DF860.
Human colon specif
Polynucleotide seq
Porphyromonas ging
Alpha-galactosidas
Tuberculous sclerosi
Human heat shock p
Human heat shock p
Human heat shock p
Shinorhizobium meli
Humanised anti-CD1
Enterococcus faeca
S. venezuelae macr

RESULT	1
ID	T72784 standard; DNA; 45 BP.
XX	
AC	T72784;
XX	
DT	22-SEP-1997 (first entry)
XX	
DE	DNA tag.
XX	
KW	Metastasis-inducing DNA; Met-DNA; cancer; diagnosis; ds.
OS	Synthetic.
XX	
FT	Key
FT	misc_feature
FT	Location/Qualifiers
FT	1..2
FT	/tag= a
FT	/note= "5' single stranded overhang"
FT	/tag= b
FT	/note= "5' overhang on complementary strand of 4
FT	bases with sequence 5'-AGCT-3'."
XX	
PN	W09725443-A1.
XX	
PD	17-JUL-1997.
XX	
PF	10-JAN-1997; 97WO-CB00074.
XX	
PR	10-JAN-1996; 96GB-0000470.
XX	
PA	(UPLI-) UNIT LIVERPOOL.
XX	
XX	Barracough BR; Rudland PS;
XX	
PI	

DR	WPI; 1997-372878/34.
XX	
PT	New isolated metastasis-inducing DNA - used to develop products to
PT	identify and treat patients at risk from metastatic tumours
XA	
PS	Claim 6; Page 24; 38pp; English.
XX	
CC	A DNA tag (T72784) having HindIII, SfiI, NotI and a defective
CC	HindIII site is used to tag human DNA from malignant, metastatic
CC	cancer cells in a novel method of screening and recovering
CC	metastasis-inducing DNA (Met-DNA). This method involves
CC	transferring the human DNA into a cell line (pref. rat Rama 37)
CC	that produces only benign, non-metastasising tumours when injected
CC	into a syngeneic animal, injecting the transformed cells into a
CC	syngenic animal, selecting those animals having metastasing
CC	tumours, and recovering the Met-DNA (see also T72785-90) from them.
CC	The human DNA is tagged to assist in its removal or insertion from
CC	or into a host or vector. The tagging procedure overcomes the
CC	problem of identifying the inserted human DNA sequences in the rat
CC	genome of the transfected cells.
XX	
SQ	Sequence 45 BP; 11 A; 13 C; 12 G; 9 T; 0 other;
Oy	
Query Match	100.0%; Score 45; DB 18; Length 45;
Best Local Similarity	100.0%; Pred. No. 7.7e-09;
Matches 45; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	
1	aatccaagcttcggcgacatcagccgaatatgcgcccattat 45
1	aatccaagcttcggcgacatcagccgaatatgcgcccattat 45
RESULT 2	
T61221	
ID	T61221 standard; DNA; 42 BP.
AC	T61221;
DT	17-OCT-1997 (first entry)
XX	
DE	Primer HUCKN.NOT for light chain constant region cDNA.
XX	
KX	Primer: polymerase chain reaction; PCR; amplification; light chain;
KM	immunoglobulin; Ig; constant; region; production;
KW	recombinant; antibody; B cell; diagnosis; therapy; ss.
OS	Synthetic.
XX	
PN	DE19526546-A1.
PD	23-JAN-1997.
PF	20-JUL-1995; 95DE-1026546.
PR	20-JUL-1995; 95DE-1026546.
PA	(OPEL/) OPELZ G.
P1	Terness P, Welschhof M;
DR	WPI; 1997-088250/09.
PT	Prodn. of recombinant antibodies - by amplification, cloning and
PT	expression of cDNA generated from B-cell mRNA
XX	
XX	Claim 5; Page 11; 11pp; German.
XX	
CC	The present sequence is a primer for the PCR amplification of the
CC	cDNA encoding an immunoglobulin (Ig), light chain, constant region.
CC	The primer can be used in a novel method for the production of
CC	recombinant antibodies, comprising the selection of B cells from a
CC	lymphocyte fraction, isolation of mRNA from individual B cells,

```

Oy      6 aagcttcgagccgatacagccgaataatgcggccg 39
      11 111111111111111111111111111111111
Db      473 AATGTTCCAGCGATCGGCGCCGAACTTGCGGCGG 440

Query Match      47.1%; Score 21.2; DB 20; Length 500;
Best Local Similarity 76.5%; Pred. No. 14;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

RESULT 3
X91235/c
ID X91235 standard; DNA; 500 BP.
XX
AC X91235;
XX
DT 24-SEP-1999 (first entry)
XX
DE T. gondii immunogenic protein encoding DNA.
XX
KW Immunogenic protein; Toxoplasma gondii protein; oocyst shedding; cat;
RV T. gondii infection; enteric apicomplexa oocyst; Cryptosporidium oocyst;
RW Toxoplasma oocyst; ss.
XX
XX Toxoplasma gondii.
OS
XX WO9932633-A1.
PN
PD 01-JUL-1999.
XX
XX 18-DEC-1998; 98WO-US27137.
PR 19-DEC-1997; 97US-0994825.
XX
PA (HESK-) HESKA CORP.
XX
PI Lutz SB, Milhausen MJ, Ng RK;
XX
DR WPI; 1999-418930/35.
XX
PT New isolated Toxoplasma gondii nucleic acids used, e.g. to treat
PT infection caused by this microorganism
XX
XX Claim 4; Page 213; 381pp; English.
XX
XX The invention provides isolated Toxoplasma gondii nucleic acids that
CC encode immunogenic polypeptides. The T. gondii nucleic acid molecules,
CC immunogenic proteins and antibodies to the proteins can be used to
CC inhibit T. gondii oocyst shedding in a cat due to infection with
CC T. gondii. They can be used for preventing T. gondii infection and for
CC preventing the spread of T. gondii infection. They can also be used for
CC detecting T. gondii infection. The detection method can be used to detect
CC parasite cysts or oocysts in feces, e.g. from enteric apicomplexa oocysts
CC such as Cryptosporidium oocysts and Toxoplasma oocysts.
XX
SQ Sequence 500 BP; 118 A; 118 C; 131 G; 133 T; 0 other;

Query Match      47.6%; Score 21.4; DB 18; Length 42;
Best Local Similarity 71.8%; Pred. No. 9.2;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy      5 caagcttcgagccgatacagccgaataatgcggccgcat 43
      |||||
Db      4 caagcttcgagccgagacagatggtgcagccacagt 42

```

```

RESULT 4
X21501/c
ID X21501 standard; DNA: 3481 BP.
XX
XX X21501:
XX
XX 21-MAY-1999 (first entry)
XX
DE DNA sequence of sorbitol dehydrogenase (SLDH) and ORF2 genes.
XX
XX Sorbitol dehydrogenase; SLDH: open reading frame; ORF2: L-sorbose;
XX D-sorbitol; vitamin C; ds.
XX
XX Gluconobacter suboxydans.
XX
XX Key Location/Qualifiers
XX RBS 177..182
XX /tag= a
XX /note= "putative Shine-Dalgarno (SD) sequence for
XX CDS 192..572
XX /tag= b
XX /product= "ORF2 gene product"
XX /note= "corresponding protein sequence shown in W95020"
XX RBS 558..564
XX /tag= c
XX /note= "putative Shine-Dalgarno (SD) sequence for
XX CDS 572..2794
XX /tag= d
XX /product= "SLDH gene product"
XX /note= "corresponding protein sequence shown in W95019"
XX sig_peptide 572..643
XX /tag= e
XX /note= "signal sequence for SLDH"
XX mat_peptide 644..2791
XX /tag= f
XX /note= "mature SLDH protein sequence"
XX repeat_region 684..704
XX /tag= g
XX /rpl_type= inverted
XX /note= "inverted repeat sequence IRI as possible
XX transcription terminator for ORF2 gene as
XX indicated in the specification"
XX repeat_unit 684..693
XX /tag= h
XX repeat_region 2803..2892
XX /tag= i
XX /rpl_type= inverted
XX /note= "inverted repeat sequence IR2 as possible
XX transcription terminator for SLDH gene as
XX indicated in the specification"
XX repeat_unit 2803..2833
XX /tag= j
XX
XX EP897984-A2.
XX
XX 24-FEB-1999.
XX
XX 13-AUG-1998; 98EP-0115231.
XX
XX 21-AUG-1997; 97EP-0114432.
XX
XX (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
XX Hosino T, Miyazaki T, Ojima S, Shinjoh M, Tomiyama N;
XX WPI: 1999-134646/12.
XX P-PSDB: W95019, W95020.
XX
XX New D-sorbitol dehydrogenase gene and recombinant protein - useful
XX

```

```

PT for production of L-sorbose, an intermediate in vitamin C production
XX
XX Claim 4; Fig 3a-D; 39pp; English.
XX
XX This represents the DNA sequence of the SLDH gene encoding a protein with
XX sorbitol dehydrogenase (SLDH) activity. The DNA also encodes an open
XX reading frame (ORF2) product upstream of the SLDH open reading frame.
XX needed for SLDH activity in vivo. Host cells transformed by a vector
XX comprising the SLDH DNA sequence are used for the recombinant expression
XX of the sorbitol dehydrogenase. Recombinant D-sorbitol dehydrogenase is
XX useful for producing L-sorbose from D-sorbitol. L-sorbose is an important
XX intermediate in vitamin C production.
XX
XX Sequence 3481 BP; 674 A; 1018 C; 1001 G; 788 T; 0 other;
XX
XX Query Match 47.1%; Score 21.2; DB 20; Length 3481;
XX Best Local Similarity 76.5%; Pred. No. 18;
XX Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
XX
XX Db 4 ccaagcttgcgagccatcagcgaatatgcgcg 37
XX ||||| ||||| ||||| ||||| |||||
XX 388 CCAAGCGTGCAGCCCATGAGCATGATATACCGCGC 355
XX
XX RESULT 5
XX ID V15518 standard; DNA: 1503 BP.
XX AC V15518;
XX
XX 22-MAY-1998 (first entry)
XX
XX Hamster oral papilloma virus L1 DNA.
XX
XX L1 DNA; L1 protein; diagnosis; papilloma virus infectious disease;
XX cancer; screening; antitumour agent; antisense treatment; ss.
XX
XX OS Hamster oral papilloma virus.
XX PN JP10042875-A.
XX
XX 17-FEB-1998.
XX
XX 06-AUG-1996; 96JP-0207143.
XX
XX 06-AUG-1996; 96GP-0207143.
XX
XX (TORA ) TORAY IND INC.
XX
XX WPI: 1998-186869/17.
XX P-PSDB: W47224.
XX
XX DNA encoding hamster oral papilloma virus protein - useful for, e.g.
XX diagnosing cancer related papilloma virus infections
XX
XX Claim 10; Pages 13-14; 17pp; Japanese.
XX
XX The present sequence is the hamster oral papilloma virus L1 DNA.
XX The L1 protein and DNA can be used for the diagnosis of papilloma
XX virus infectious diseases related to cancer, screening for
XX antitumour agents and antisense treatment.
XX
XX Sequence 1503 BP; 434 A; 346 C; 337 G; 386 T; 0 other;
XX
XX Query Match 46.2%; Score 20.8; DB 19; Length 1503;
XX Best Local Similarity 70.0%; Pred. No. 23;
XX Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
XX
XX Db 6 aagcttgcgagccatcagcgaatatgcgcgcatat 45
XX ||||| ||||| ||||| ||||| |||||
XX 826 aaactctgcgctcagatcagatcaaaagctgatcatcat 865
XX

```

```

RESULT 6
V15519 ID V15519 standard; DNA: 7647 BP.
XX
AC V15519;
XX
DT 22-MAY-1998 (first entry)
XX
DE Hamster oral papilloma virus genomic DNA.
XX
KW Diagnosis; papilloma virus infectious disease;
XX cancer; screening; antitumour agent; antisense treatment; ss.
OS Hamster oral papilloma virus.
XX
PM JP10042875-A.
XX
PD 17-FEB-1998.
XX
PF 06-AUG-1996; 96JP-0207143.
XX
PR 06-AUG-1996; 96JP-0207143.
XX
PA (TORA ) TORAY IND INC.
XX
DR WPI: 1998-186869/17.
XX
PT DNA encoding hamster oral papilloma virus protein - useful for, e.g.
XX diagnosing cancer related papilloma virus infections
XX
PS Claim 11; Pages 14-16; 17pp: Japanese.
XX
CC The present sequence is a hamster oral papilloma virus genomic
CC DNA. The DNA can be used for the diagnosis of papilloma virus
CC infectious diseases related to cancer; screening for antitumour
CC agents and antisense treatment.
XX
SQ Sequence 7647 BP; 2102 A; 1703 C; 1894 G; 1948 T; 0 other;

Query Match 46.2%; Score 20.8; DB 19; Length 7647;
Best Local Similarity 70.0%; Pred. No. 28;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 6 aagcttgccgcatcagccgaatatgcgcgcattat 45
   ||| ||||| ||||| ||| ||| ||||| ||
DB 6474 aaactcgcgctgcatcagatcaaaagctgcatcat 6513

RESULT 7
236926 ID 236926 standard; DNA: 8907 BP.
XX
AC 236926;
XX
DT 13-MAR-2000 (first entry)
XX
DE Nucleotide sequence of the genome of Stealth virus clone 313.
XX
KW Prototype stealth virus clone; atypically structured virus;
KW vacuolating cytopathic effect; cytopathic virus; virus detection;
KW malignancy; multiple myeloma; lymphoma; brain tumour; breast cancer;
KW salivary gland tumour; Alzheimer's disease; Parkinson's disease;
KW spongiform encephalopathy; multiple sclerosis; schizophrenia;
KW manic depression; major depression; personality disorder; autism;
KW Rett's syndrome; attention deficit; oppositional defiance; aggression;
KW anorexia nervosa; bulimia; multi-system illness; virus infection; ss.
XX
OS Stealth virus.
XX
PM WO9960101-A1.

```

```

XX
PD 25-NOV-1999.
XX
PF 19-MAY-1999; 99WO-US11185.
XX
PR 19-MAY-1998; 98US-0081708.
XX
PA (MART/) MARTIN W J.
XX
PI Martin WJ;
XX
DR WPI: 2000-072436/06.
XX
PT Isolated viruses, used to develop products for detection, prevention
XX and treatment of stealth virus infections
XX
PS Disclosure; Page 34-37; 50pp: English.
XX
CC 236925-30 represent the nucleotide sequences of prototype Stealth virus
CC clones. The sequences illustrate the type of sequences that can be used
CC to generate probes and to predict antigenic and biologically active
CC products obtainable from a viral isolate, in the methods of the
CC invention. The specification describes tissue culture, serological and
CC molecular based methods to detect atypically structured viruses,
CC such as Stealth viruses, which are able to induce a vacuolating
CC cytopathic effect (CPE) in tissue culture, and are distinguishable
CC from known cytopathic viruses, by appearance progression and/or host
CC range, or by serological, electron microscopic and/or molecular markers.
CC The products can be used for detecting viruses in patients with
CC diseases such as a malignancy, e.g. multiple myeloma, lymphoma, brain
CC tumours, breast cancer, salivary gland tumours, Alzheimer's disease,
CC Parkinson's disease, spongiform encephalopathy, multiple sclerosis,
CC schizophrenia, manic depression, major depression, personality
CC disorders, autism, Rett's syndrome, attention deficit, oppositional
CC defiance, aggression, anorexia nervosa, bulimia, a multi-system
CC illness, an animal illness or an illness in a domestic dog or cat.
CC The products can be used for detecting Stealth viruses in biological
CC products such as blood products and foods. The products can also be
CC used for developing agents for treating or preventing virus infections.
XX
SQ Sequence 8907 BP; 2257 A; 2447 C; 2165 G; 2023 T; 15 other;

Query Match 46.2%; Score 20.8; DB 21; Length 8907;
Best Local Similarity 73.5%; Pred. No. 28;
Matches 25; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 5 caagcttgccgcatcagccgaatatgcgcgc 38
   ||| ||||| ||||| ||| ||| ||||| ||
DB 6014 caccatcgggcgccgcatcagccgaccatcagcc 6047

RESULT 8
X84323 ID X84323 standard; DNA: 9181 BP.
XX
AC X84323;
XX
DT 08-SEP-1999 (first entry)
XX
DE Stealth virus nucleic acid clone, SEQ ID NO: 15.
XX
KW Stealth virus; detection; diagnosis; infection; ss.
XX
OS Stealth virus.
XX
FH Key Location/Qualifiers
FT misc_difference 8920
FT /tag= a
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the
FT base numbering given in the specification"
FT
FT misc_difference 8929

```

[illegible]

Query Match	46.2%	Score 20.8:	DB 20:	Length 9181:
Best Local Similarity	73.5%:	Pred. No. 28:		
Matches 25:	Conservative 1:	Mismatches 8:	Indels 0:	Gaps 0:
5 caagcttgcggccgcatcagccgaatcgtcgcc	38			

Db 6020 caccatcgggccgatcagcgaccatrrcagcc 6053

RESULT 9

ID Q701.14 standard; DNA; 41 BP.

AC Q70114;

DT 10-MAR-1995 (first entry)

DE Lambda INNER primer.

KW Enterokinase; EK; heavy chain; light chain; catalytic domain;

KW trypsin; enzyme; PACE gene; ss.
 YX

OS Synthetic.

PN W09416083-A-
XY

PD 21-JUL-1994

PF 13-JAN-1994; 94WO-0500616.
XX

PK 15-JAN-1993; 9305-0005944.
XX

PA (GEM) GENETICS INST INC.
XX

PL Lavalley ER;
XX

WFL; 1994-249229/30.
XX

PT vectors, host cells, expression products and antibodies are

PT proteins used in creating insoluble and for creating fusion

PS Example 6; Page 19; 50pp; English.

Nested oligonucleotide primers were synthesised which were

complementarily to the lambda DNA sequence adjacent to the cloning site for the cDNA insertions (see 070113-14). In addition, primers were

region of the EK coding sequence (see Q70115-16). Subcloning and

CC contained regions corresp. to tryptic peptides (see R57295-9).

Sequence 41 BP; 12 A; 10 C; 11 G; 8 T; 0 other;

Query Match	45.38; Score 20.4; DB 15; Length 41;
-------------	--------------------------------------

Matches	27;	Conservative	0;	Mismatches	11;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

QY 8 gcttgcgcgcgatcagggccgaatatatgcggccgcatlta 45

Db 41 GCTTGACTTAACCAAGCTGACTTGGGGCCGATTCT 4

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Z29251
ID: Z30351 at 2023-03-23. DNA: 030 BB

XX
AC 729251.

XX	28-FEB-2000 (first entry)
DT	

XX Rifin 3193 gene
DE

XX Rif1n 3193 gene: P falciarum chromosome 2: Rif1n protein:

an influenza expression plasmid VR1050; DNA vaccine; immunisation;

KM stage specific protein expression; antimicrobial vaccine;
 KM antimicrobial drug; IAF; immunofluorescent antibody testing; ds.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO9955381-A1.
 XX
 PD 04-NOV-1999.
 XX
 PF 26-APR-1999; 99WO-US09047.
 XX
 PR 24-APR-1998; 98US-0082947.
 PR 23-APR-1999; 99US-0082947.
 XX
 PA (USNA) US SEC OF NAVY.
 XX
 PI Hoffman SL, Carucci DJ;
 DR WPI; 2000-086380/07.
 XX
 PT use of microbial, animal and/or human genomic data for identification
 of vaccine, drug or diagnostic products -
 XX
 PS Claim 13; Page 16; 23pp; English.
 XX
 CC The present sequence is the Rifin 3193 gene of chromosome 2. Rifin genes
 were used to characterise the protein expression from various life cycle
 stages of P.falciparum. Oligonucleotide primers were used to amplify each
 selected ORF from chromosome 2 using genomic DNA as template. The
 CC purified amplified products were cloned into the mammalian expression
 CC plasmid VR1050 to produce DNA vaccines for immunisation. Blood and sera
 obtained from groups of mice immunised with doses of vaccine was used to
 CC identify stage specific expression of Rifin protein by immunofluorescent
 CC antibody testing. The method is useful for determination of subcellular
 CC localisation of proteins and for the development of antimicrobial
 CC vaccines and drugs.
 SQ Sequence 939 BP; 365 A; 133 C; 172 G; 269 T; 0 other;

Query Match 45.3%; Score 20.4; DB 21; Length 939;
 Best Local Similarity 71.1%; Pred. No. 31;
 Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 atccaagcttgccgagccgaatgagccg 39
 DB 702 atccaagcgagctgcgcatgagatgagccg 739

RESULT 11
 N81768/C
 ID N81768 standard; DNA: 4260 BP.
 XX
 AC N81768;
 XX
 DE 29-DEC-1990 (first entry)
 XX
 DE Sequence encoding Mycobacterium tuberculosis 540 and 517 AA residue
 DE proteins.
 XX
 KM Diagnosis; assay; M.bovis; vaccine; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN Key Location/Qualifiers
 FH CDS 252..1874
 FT /*tag= a
 FT /label=540 AA protein
 FT /note="P81351"
 FT complement (3948..2395)
 FT /*tag= b
 FT /label=517 AA protein
 FT /note="P81868"

XX
 PN MO8806591-A.
 XX
 PD 07-SEP-1988.
 XX
 PF 25-FEB-1988; 88WO-US00598.
 XX
 PR 24-FEB-1988; 88US-0159667.
 PR 06-FEB-1987; 87US-0019529.
 XX
 PA (SCRI-) SCRIPPS CLINIC & RE.
 XX
 PI Shinnick T, Houghten R;
 DR WPI; 1988-271136/38.
 DR P-PSDB; P81351, P81868.
 XX
 PT Recombinant mycobacterial peptide(s) -
 PT used in assays for diagnosis of infection, for producing
 PT vaccines and for producing antibodies
 XX
 PS Disclosure; Fig 2a-2d; 116pp; English.
 XX
 CC An isolated DNA molecule that consists essentially of the nucleotide
 CC sequence that corresponds to the sequence represented by position 3950 to
 CC about 2390 and from position 3948 through position 2398 of N81768 is
 CC claimed. Also claimed is a peptide sequence that consists of a 5-40 AA
 CC residue sequence that corresponds to a sequence of the 540 AA residue
 CC protein (P81351) or the 517 AA residue protein (P81868) coded for by the
 CC DNA sequence. The proteins can be used for determining previous
 CC immunological exposure of a mammal to M.tuberculosis or M.bovis and
 CC for producing a vaccine.
 SQ Sequence 4260 BP; 733 A; 1332 C; 1481 G; 714 T; 0 other;

Query Match 45.3%; Score 20.4; DB 9; Length 4260;
 Best Local Similarity 80.0%; Pred. No. 37;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 12 gcggcgatcagcgccgaatgagccgca 41
 DB 3510 GCGGCAAGCGAGCGGAAACAGCGCCGCA 3481

RESULT 12
 N80222/C
 ID N80222 standard; DNA: 4380 BP.
 XX
 AC N80222;
 XX
 DE 19-MAR-1991 (first entry)
 XX
 DE Sequence of Mycobacterium tuberculosis DNA contg. gene encoding 65
 DE protein.
 XX
 KM Antigen; vaccine; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN Key Location/Qualifiers
 FH CDS 192..1874
 FT /*tag= a
 FT complement (2398..4101)
 FT /*tag= b
 FT CDS
 XX
 PN WO8805823-A.
 XX
 PD 11-AUG-1988.
 XX
 PR 01-FEB-1988; 88WO-US00281.
 PR 02-FEB-1987; 87US-0010007.

XX (WHT-) WHITEHEAD INST BIOM.
XX
XX
PI Hussion RN, Young RA, Shinnick TW.
XX
DR WPI: 1988-235175/33.
XX P-PSDB: P80215, P80216
XX
PT Genes encoding Mycobacterium tuberculosis protein antigens -
PT useful for developing reagents for diagnosis, prevention and
PT treatment of tuberculosis
XX
PS Claim 12; Fig 8; 82pp; English.
XX
CC The gene was isolated by probing a lambda gtl1 expression library of
CC M.tuberculosis DNA with monoclonal antibodies directed against
CC M.tuberculosis-specific antigens. The 19KD, 71KD and the 65KD proteins
CC and genes are claimed, and so is a vaccine comprising DNA encoding
CC M.tuberculosis protein in a recombinant vaccine vector. P80216 is
CC encoded on the complementary strand.
XX
SQ Sequence 4380 BP; 757 A; 1373 C; 1512 G; 738 T; 0 other;

Query Match 45.3%; Score 20.4; DB 9; Length 4380;
Best Local Similarity 80.0%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 12 gcggcgatcagccgatatgcgcgca 41
DB 3630 GCGGCAAGCAGCGGCAACAGCGCGCA 3601

RESULT 13
V05708/c
ID V05708 standard; DNA; 4380 BP.
XX
AC V05708;
XX
DT 22-JUN-1998 (first entry)
XX
DE Mycobacterium tuberculosis 65 kda heat shock protein gene.
XX
KW Heat shock protein; Mt Hsp65; autoimmune disease; immunotherapy;
KW gene therapy; rheumatoid arthritis; multiple sclerosis; ds.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT 252..1874
FT /*tag= a
FT /product= 65 kda heat shock protein
XX
PN W09746253-A2.
XX
XX 11-DEC-1997.
XX
PD 03-JUN-1997; 97WO-US09427.
XX
PE 03-JUN-1997; 97US-0019100.
XX
PR 03-JUN-1997; 96US-0019100.
XX
XX (AURA-) AURAGEN INC.
XX
PI Haynes JR, Prayaga SK, Ramshaw IA;
XX WPI: 1998-041892/04.
XX P-PSDB: W44702.
XX
XX Treatment of autoimmune diseases - by administering
XX autoantigen-coated particles or autoantigen-encoding nucleic acid
XX construct

PS Example 2; Page 55-59; 72pp; English.
XX
XX This DNA sequence encodes the 65 kda heat shock protein (see
CC W44702), designated Mt Hsp65, of Mycobacterium tuberculosis. This
CC protein cross-reacts with a component of articular cartilage, human
CC Hsp60, that is up-regulated in the joints of arthritic patients. A
CC claimed method for treating or preventing an autoimmune disease in
CC a mammal comprises: (a) providing a particle coated with an antigen
CC against which an immune response is mounted in the autoimmune
CC disease; (b) delivering the particle into the recipient cell of the
CC mammal; and (c) repeating step (b) until either a reduction in a
CC cytotoxic immune response or a desensitizing immune response is
CC induced in the mammal. Alternatively, step (a) comprises providing
CC a nucleic acid construct comprising a coding sequence for the
CC antigen, operably linked to control elements such that the coding
CC sequence can be transcribed and translated in a recipient cell, and
CC delivering the construct to the recipient cell using a gene gun.
CC The antigen of step (a) is selected from collagen, Mt Hsp65,
CC myelin basic protein, myelin oligodendrocyte glycoprotein,
CC proteolipid protein, and epitopes thereof. These antigens mitigate
CC cytotoxic responses and elicit antigen desensitisation. The method
CC is used especially for treating rheumatoid arthritis or multiple
CC sclerosis. It represents a novel use for the known Mt Hsp65 gene.
XX
SQ Sequence 4380 BP; 757 A; 1371 C; 1514 G; 738 T; 0 other;

Query Match 45.3%; Score 20.4; DB 19; Length 4380;
Best Local Similarity 80.0%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 12 gcggcgatcagccgatatgcgcgca 41
DB 3630 GCGGCAAGCAGCGGCAACAGCGCGCA 3601

RESULT 14
Z46842/c
ID Z46842 standard; CDNA; 1818 BP.
XX
AC Z46842;
XX
DT 10-APR-2000 (first entry)
XX
DE Semaphorin K1 gene related sequence.
XX
KW Semaphorin K1; cellular physiology; neurite outgrowth; neuron; human;
KW Immunogen; pharmaceutical; ss.
XX
XX Unidentified.
XX
OS JP11341988-A.
XX
PN 14-DEC-1999.
XX
PD 11-MAR-1999; 99JP-0005672.
XX
PE 11-MAR-1998; 98US-0041236.
XX
PR (EXEL-) EXELIXIS PHARM INC.
XX
XX WPI: 2000-109378/10.
XX P-PSDB: Y56855.
XX
XX New semaphorin polypeptides, useful cell physiology modulators and
XX immunogens -
XX
PS Disclosure; Page 17-20; 57pp; Japanese.
XX
XX The invention provided isolated human semaphorin K1 polypeptides. The
XX polypeptides, or nucleic acids encoding them, can be used to modulate
XX cellular physiology by modulating semaphorin K1 activity, e.g. semaphorin
XX K1 polypeptide fragments or antisense nucleic acids can be used to

Mon Mar 5 13:21:05 2001

us-09-101-423b-7.rng
